

Eself_case_study_1_IEEE_fraud_detection_EDA

Overview

1. Many institutions are being victimized by fraudulent transactions, which costs them millions of dollars. Machine learning models can be used to identify fraud triggers and create trackers. These trackers can detect such frauds and provide real-time mitigation assistance.
2. The job at hand is to predict fraudulent transactions in real time and decline them using the provided data. So, it is our task to create a predictive model that can identify and stop such fraud transactions before they happen. The difficult aspect of the issue is that there are remarkably few fraud cases in the dataset (based on observation) when compared to the population as a whole.
3. This is a binary classification exercise, where "1" denotes a fraudulent transaction and "0" a legitimate one.

Data Source

The data comes from Vesta's real-world e-commerce transactions and contains a wide range of features such as device type, product features, transaction card details, transaction amount, personal information such as email address etc. [CLICK HERE FOR DATA](#)

Dataset Description

The datasets provided by the Competiton Host are as follows:

train_transaction.csv : The transaction dataset comprising the transaction details to be used for training the model.

train_identity.csv : The identity dataset comprising the additional details about the identity of the payer and the merchant between whom the transaction was performed and the details of transactions are present in the train_transaction.csv.

test_transaction.csv : The transaction dataset comprising the transaction information to test the performance of the trained model.

test_identity.csv : The identity dataset comprising the additional identity information about the transactions present in the test_transaction.

In this competition you are predicting the probability that an online transaction is fraudulent, as denoted by the binary target **isFraud**.

Description of Transaction Dataset

TransactionID: Id of the transaction and is the foreign key in the Identity Dataset.

isFraud: 0 or 1 signifying whether a transaction is fraudulent or not.

TransactionDT: timedelta from a given reference datetime (not an actual timestamp)

TransactionAMT: Transaction Payment Amount in USD.

ProductCD: Product Code.

addr: Address

dist: Distance

P_emaildomain: Purchaser Email Domain.

R_emaildomain: Receiver Email Domain.

C1-C14: counting, such as how many addresses are found to be associated with the payment card, etc. The actual meaning is masked.

D1-D15: timedelta, such as days between previous transactions, etc.

M1-M9: match, such as names on card and address, etc.

V1-V339: Vesta engineered rich features, including ranking, counting, and other entity relations.

Following Features are Categorical in the Transaction Dataset,

ProductCD

card1 - card6

addr1, addr2

P_emaildomain

R_emaildomain

M1-M9

Description of the Identity Dataset

Following Features are present in the Identity Dataset,

TransactionID: Foreign key to the Transaction Dataset.

id_01-id_38: Masked features corresponding to the identity of the card holders.

DeviceType: Type of Device used to make the Transaction.

DeviceInfo: Information regarding the characteristics of the Device.

Following Features are Categorical in the Identity Dataset,

DeviceType

DeviceInfo

id_12 — id_38

Business Problem

We will be predicting whether a transaction is fraudulent or not and to predict this we will be building an ML model using a real-world e-commerce dataset. The goal is to improve the efficacy of fraudulent transaction alerts for millions of people around the world, helping hundreds of thousands of businesses reduce their fraud loss and increase their revenue.

Exploratory Data Analysis

```
In [1]: # importing the libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import random
from collections import Counter
from prettytable import PrettyTable
from time import time
import datetime

%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
```



```
In [2]: # reading the datasets
df_train_transaction = pd.read_csv("ieee-fraud-detection/train_transaction.csv")
df_train_identity = pd.read_csv("ieee-fraud-detection/train_identity.csv")
df_test_transaction = pd.read_csv("ieee-fraud-detection/test_transaction.csv")
df_test_identity = pd.read_csv("ieee-fraud-detection/test_identity.csv")
```



```
In [3]: #getting the shape of each train dataset and test dataset
print("shape of train_transaction dataset:", df_train_transaction.shape)
print("shape of train_identity dataset:", df_train_identity.shape)
print("shape of test_transaction dataset:", df_test_transaction.shape)
print("shape of test_identity dataset:", df_test_identity.shape)
```



```
shape of train_transaction dataset: (590540, 394)
shape of train_identity dataset: (144233, 41)
shape of test_transaction dataset: (506691, 393)
shape of test_identity dataset: (141907, 41)
```



```
In [4]: # merging transaction and identity data
df_train = df_train_transaction.merge(df_train_identity, on = 'TransactionID', how = 'left')
target = df_train['isFraud']      #storing class label inside a 'target' variable
# df_train.drop(['isFraud'], axis=1, inplace = True)
```

```
df_test = df_test_transaction.merge(df_test_identity, on = 'TransactionID', how =
```

```
In [5]: print(df_train.head(4))
print('*'*80)
print(df_test.head(4))
```

	TransactionID	isFraud	TransactionDT	TransactionAmt	ProductCD	card1	\
0	2987000	0	86400	68.5	W	13926	
1	2987001	0	86401	29.0	W	2755	
2	2987002	0	86469	59.0	W	4663	
3	2987003	0	86499	50.0	W	18132	

	card2	card3	card4	card5	...	id_31	id_32	id_33	id_34	id_35	\
0	NaN	150.0	discover	142.0	...	NaN	NaN	NaN	NaN	NaN	
1	404.0	150.0	mastercard	102.0	...	NaN	NaN	NaN	NaN	NaN	
2	490.0	150.0	visa	166.0	...	NaN	NaN	NaN	NaN	NaN	
3	567.0	150.0	mastercard	117.0	...	NaN	NaN	NaN	NaN	NaN	

	id_36	id_37	id_38	DeviceType	DeviceInfo
0	NaN	NaN	Nan	NaN	NaN
1	NaN	NaN	Nan	NaN	NaN
2	NaN	NaN	Nan	NaN	NaN
3	NaN	NaN	Nan	NaN	NaN

[4 rows x 434 columns]

	TransactionID	TransactionDT	TransactionAmt	ProductCD	card1	card2	\
0	3663549	18403224	31.95	W	10409	111.0	
1	3663550	18403263	49.00	W	4272	111.0	
2	3663551	18403310	171.00	W	4476	574.0	
3	3663552	18403310	284.95	W	10989	360.0	

	card3	card4	card5	card6	...	id-31	id-32	id-33	id-34	id-35	id-36	\
0	150.0	visa	226.0	debit	...	NaN	NaN	NaN	NaN	NaN	NaN	
1	150.0	visa	226.0	debit	...	NaN	NaN	NaN	NaN	NaN	NaN	
2	150.0	visa	226.0	debit	...	NaN	NaN	NaN	NaN	NaN	NaN	
3	150.0	visa	166.0	debit	...	NaN	NaN	NaN	NaN	NaN	NaN	

	id-37	id-38	DeviceType	DeviceInfo
0	NaN	NaN	Nan	NaN
1	NaN	NaN	Nan	NaN
2	NaN	NaN	Nan	NaN
3	NaN	NaN	Nan	NaN

[4 rows x 433 columns]

```
In [6]: print(df_train.columns)
print('*'*80)
print(df_test.columns)
```

```
Index(['TransactionID', 'isFraud', 'TransactionDT', 'TransactionAmt',
       'ProductCD', 'card1', 'card2', 'card3', 'card4', 'card5',
       ...
       'id_31', 'id_32', 'id_33', 'id_34', 'id_35', 'id_36', 'id_37', 'id_38',
       'DeviceType', 'DeviceInfo'],
      dtype='object', length=434)
```

```
=====
Index(['TransactionID', 'TransactionDT', 'TransactionAmt', 'ProductCD',
       'card1', 'card2', 'card3', 'card4', 'card5', 'card6',
       ...
       'id-31', 'id-32', 'id-33', 'id-34', 'id-35', 'id-36', 'id-37', 'id-38',
       'DeviceType', 'DeviceInfo'],
      dtype='object', length=433)
```

We observed that The names of the id features in test dataset and train dataset did not match. While id features in the test dataset were of the form id-x, they were present in the train dataset with the name id_x, where x was a number between 01 and 38. and as a result, we will convert the test dataset's id feature names from id-x to id_x.

```
In [7]: # converting the test dataset's id feature names from id-x to id_x
for i in df_test.columns:
    k = i.replace('-', '_')
    df_test.rename(columns = {i:k}, inplace = True)
```

```
In [8]: print(df_train.columns)
print('='*50)
print(df_test.columns)

Index(['TransactionID', 'isFraud', 'TransactionDT', 'TransactionAmt',
       'ProductCD', 'card1', 'card2', 'card3', 'card4', 'card5',
       ...
       'id_31', 'id_32', 'id_33', 'id_34', 'id_35', 'id_36', 'id_37', 'id_38',
       'DeviceType', 'DeviceInfo'],
      dtype='object', length=434)
=====
Index(['TransactionID', 'TransactionDT', 'TransactionAmt', 'ProductCD',
       'card1', 'card2', 'card3', 'card4', 'card5', 'card6',
       ...
       'id_31', 'id_32', 'id_33', 'id_34', 'id_35', 'id_36', 'id_37', 'id_38',
       'DeviceType', 'DeviceInfo'],
      dtype='object', length=433)
```

```
In [9]: # df_train.describe()
```

```
In [10]: # getting the missing data from train dataset
total = df_train.isnull().sum().sort_values(ascending=False)
percent = (df_train.isnull().sum()/df_train.isnull().count()).sort_values(ascending=False)
missing_data_train = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing_data_train.head(10)
```

	Total	Percent
id_24	585793	0.991962
id_25	585408	0.991310
id_07	585385	0.991271
id_08	585385	0.991271
id_21	585381	0.991264
id_26	585377	0.991257
id_27	585371	0.991247
id_23	585371	0.991247
id_22	585371	0.991247
dist2	552913	0.936284

```
In [11]: # getting the missing data from test dataset
total = df_test.isnull().sum().sort_values(ascending=False)
percent = (df_test.isnull().sum()/df_test.isnull().count()).sort_values(ascending=False)
missing_data_test = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing_data_test.head(10)
```

	Total	Percent
id_24	501951	0.990645
id_25	501652	0.990055
id_26	501644	0.990039
id_21	501632	0.990016
id_08	501632	0.990016
id_07	501632	0.990016
id_27	501629	0.990010
id_23	501629	0.990010
id_22	501629	0.990010
dist2	470255	0.928090

```
In [12]: # Observation: Train and test both the datasets are having high number of missing values
```

```
In [13]: ## duplicates in dataset
print("number of duplicate rows in train dataset:", df_train.duplicated().sum())
print("number of duplicate rows in test dataset:", df_test.duplicated().sum())
```

number of duplicate rows in train dataset: 0
 number of duplicate rows in test dataset: 0

Observation: There are no duplicate rows in our dataset

1.0 Univariate Analysis

1.1 Univariate Analysis of categorical features

```
In [14]: # available categorical features based on information provided
cat_fea = ['ProductCD', 'card1', 'card2', 'card3', 'card4', 'card5', 'card6', 'addr1',
           'P_emaildomain', 'R_emaildomain', 'M1', 'M2', 'M3', 'M4', 'M5', 'M6', 'M7',
           'DeviceType', 'DeviceInfo', 'id_12', 'id_13', 'id_14', 'id_15', 'id_16',
           'id_19', 'id_20', 'id_21', 'id_22', 'id_23', 'id_24', 'id_25', 'id_26',
           'id_29', 'id_30', 'id_31', 'id_32', 'id_33', 'id_34', 'id_35', 'id_36']
```

```
In [15]: def cat_feature_analysis(df_train, feature, target):

    colour = ["#" + ''.join([random.choice('0123456789ABCDEF') for j in range(6)]) for i in range(10)]

    temp = pd.DataFrame(df_train.groupby(feature)[target].apply(np.mean)).reset_index()
    temp.columns = [feature, '% Fraud Transactions']
    temp.sort_values(by=['% Fraud Transactions'], ascending = False, inplace=True)
    plt.figure(figsize=(20,5))
    plt.bar(temp[feature], temp['% Fraud Transactions'], color = random.choice(colour))
    plt.xlabel(feature)
    plt.ylabel("% rate of fraud transactions")
    plt.xticks(rotation = 90)
    plt.show()
```

```
In [16]: #stacked bar plots matplotlib: https://matplotlib.org/gallery/lines_bars_and_markers/stacked_bars.html
def stack_plot(data, xtick, col2='isFraud', col3='total'):
    ind = np.arange(data.shape[0])

    plt.figure(figsize=(20,5))
    p1 = plt.bar(ind, data[col3].values)
    p2 = plt.bar(ind, data[col2].values)

    plt.ylabel('Observations')
    plt.title('Transactions')
    plt.xticks(ind, list(data[xtick].values), rotation = 90)
    plt.legend((p1[0], p2[0]), ('Total', 'Fraud Transactions'))
    plt.show()
```

```
In [17]: def univariate_barplots(data, col1, col2='isFraud'):
    # Count number of zeros in datafram python: https://stackoverflow.com/a/5154053
    temp = pd.DataFrame(data.groupby(col1)[col2].agg(lambda x: x.eq(1).sum())).reset_index()

    # Pandas dataframe groupby count: https://stackoverflow.com/a/19385591/4084039
    temp['total'] = pd.DataFrame(data.groupby(col1)[col2].agg(total='count')).reset_index()
    temp['% Occurrence'] = temp['total'] / data.shape[0]
    temp['Avg Fraud'] = pd.DataFrame(data.groupby(col1)[col2].agg(Avg='mean')).reset_index()

    temp.sort_values(by=['total'], inplace=True, ascending=False)

    stack_plot(temp, xtick=col1, col2=col2, col3='total')
    print('\n')
    print(temp)
```

```
In [18]: def plot_categorical_features(data, feature, target):

    cat_fea = ['ProductCD', 'card1', 'card2', 'card3', 'card4', 'card5', 'card6',
               'P_emaildomain', 'R_emaildomain', 'M1', 'M2', 'M3', 'M4', 'M5', 'M6', 'M7',
               'DeviceType', 'DeviceInfo', 'id_12', 'id_13', 'id_14', 'id_15', 'id_16',
               'id_17', 'id_18', 'id_19', 'id_20', 'id_21', 'id_22', 'id_23', 'id_24', 'id_25', 'id_26',
               'id_27', 'id_28', 'id_29', 'id_30', 'id_31', 'id_32', 'id_33', 'id_34', 'id_35', 'id_36']

    if feature in cat_fea:

        data_legit = data.loc[data[target]== 0]
        data_fraud = data.loc[data[target]== 1]

        x = PrettyTable()
        x.field_names = ["Feature Description", "Feature Description of legit transactions", "Feature Description of fraud transactions"]
        x.add_row([data[feature].describe(), data_legit[feature].describe(), data_fraud[feature].describe()])
        print('\n')
        print('*Five Number Summary*')
        print(x)
        print('\n')
        print(f'The unique values in {feature} feature is:', data[feature].nunique())
        print(f'{feature} feature has {data[feature].isnull().sum()} null values,')
        print('\n')
        cat_feature_analysis(data, feature, target)
        univariate_barplots(data, feature, target)

        if (data[feature].dtype == 'int64') or (data[feature].dtype == float):
            plt.figure(figsize=(12,6))
            plt.subplot(1,2,1)
            sns.distplot(data_legit[feature], label = 'legit transactions')
            sns.distplot(data_fraud[feature], label = 'fraud transactions')
            plt.legend()
```

```

plt.title(f'Analysis on {feature} feature')
plt.grid()
plt.subplot(1,2,2)
sns.histplot(x=data[feature], hue=target, kde= True, data=data)
plt.title(f'Analysis on {feature} feature')
plt.grid()
plt.show()

```

1.1.1 target variable ('isFraud')

In [19]:

```

# this code is taken from
# https://matplotlib.org/gallery/pie_and_polar_charts/pie_and_donut_labels.html#spf

y_value_counts = target.value_counts()
print("Number of fraud transactions ", y_value_counts[1], ", (", (y_value_counts[1]
print("Number of legitimate transactions ", y_value_counts[0], ", (", (y_value_counts[0]

fig, ax = plt.subplots(figsize=(6, 6), subplot_kw=dict(aspect="equal"))
recipe = ["Fraud Transaction", "Legitimate Transaction"]

data = [y_value_counts[1], y_value_counts[0]]

wedges, texts = ax.pie(data, wedgeprops=dict(width=0.5), startangle=-40)

bbox_props = dict(boxstyle="square,pad=0.3", fc="w", ec="k", lw=0.72)
kw = dict(xycoords='data', textcoords='data', arrowprops=dict(arrowsize=1,
    bbox=bbox_props, zorder=0, va="center"))

for i, p in enumerate(wedges):
    ang = (p.theta2 - p.theta1)/2. + p.theta1
    y = np.sin(np.deg2rad(ang))
    x = np.cos(np.deg2rad(ang))
    horizontalalignment = {-1: "right", 1: "left"}[int(np.sign(x))]
    connectionstyle = "angle,angleA=0,angleB={}".format(ang)
    kw["arrowprops"].update({"connectionstyle": connectionstyle})
    ax.annotate(recipe[i], xy=(x, y), xytext=(1.35*np.sign(x), 1.4*y),
        horizontalalignment=horizontalalignment, **kw)

ax.set_title("Number of Fraud Transaction against Legitimate Transaction")

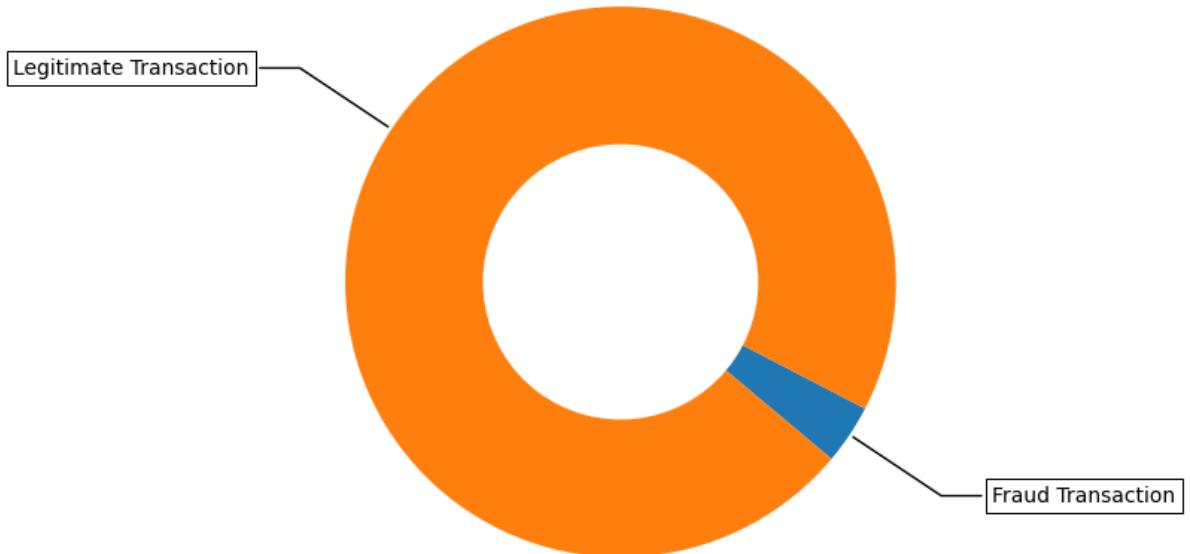
plt.show()

```

Number of fraud transactions 20663 , (3.4990009144173126 %)

Number of legitimate transactions 569877 , (96.50099908558268 %)

Number of Fraud Transaction against Legitimate Transaction



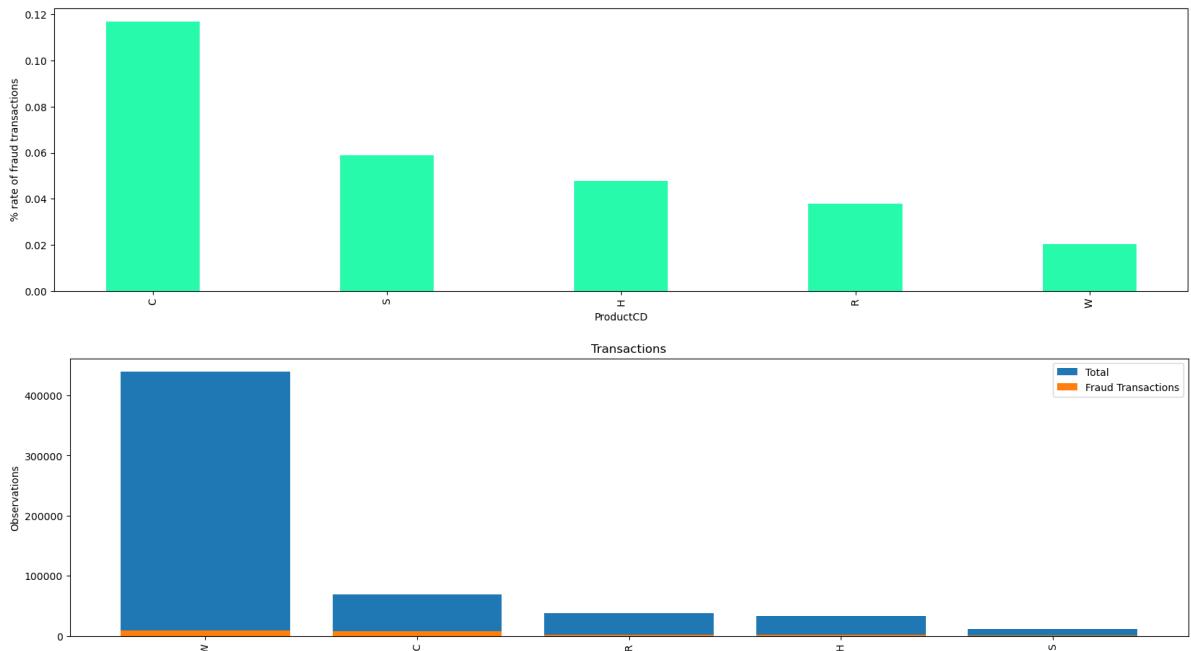
Observation: Number of fraud transaction are very low (3.5%) as compared to legitimate transactions(96.5%). Therefore, we are dealing with an extremely imbalanced dataset.

1.1.2 Categorical Feature: ProductCD

```
In [20]: plot_categorical_features(df_train, 'ProductCD', 'isFraud')
```

Five Number Summary					
		Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
		count		count	
count	20663		590540		569877
unique	5		5		5
top	W		W		W
freq	439670		8969		430701
Name: ProductCD, dtype: object			Name: ProductCD, dtype: object		
Name: ProductCD, dtype: object					

The unique values in ProductCD feature is: 5
 ProductCD feature has 0 null values, which holds 0.0 %



	ProductCD	isFraud	total	% Occurrence	Avg Fraud
4	W	0	8969	439670	0.744522
0	C	1	8008	68519	0.116028
2	R	0	1426	37699	0.063838
1	H	1	1574	33024	0.055922
3	S	0	686	11628	0.019690

Observation: There are a total of 5 categories in the feature "ProductCD" ("W", "C", "R", "H", and "S"). The category that appears the most frequently in this is 'W'. It occurs at 74.45%, however it is the most secure category with a 2% probability of being identified as a fraudulent transaction. Whereas category "C" occurs at 11.68%. We discovered that there is an 11.68% likelihood of a fraud transaction occurring when a product falls into the "C" category which is the most by any category.

1.1.3 Card Features

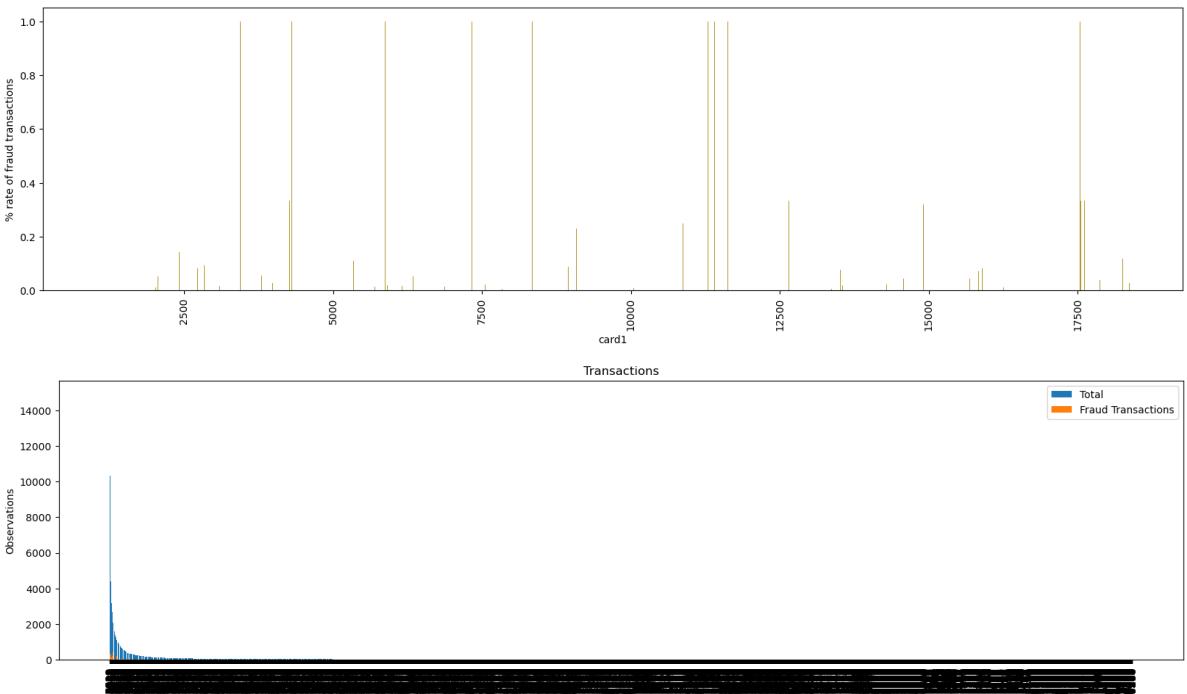
1.1.3.1 Card1

```
In [21]: plot_categorical_features(df_train, 'card1', 'isFraud')
```

Five Number Summary

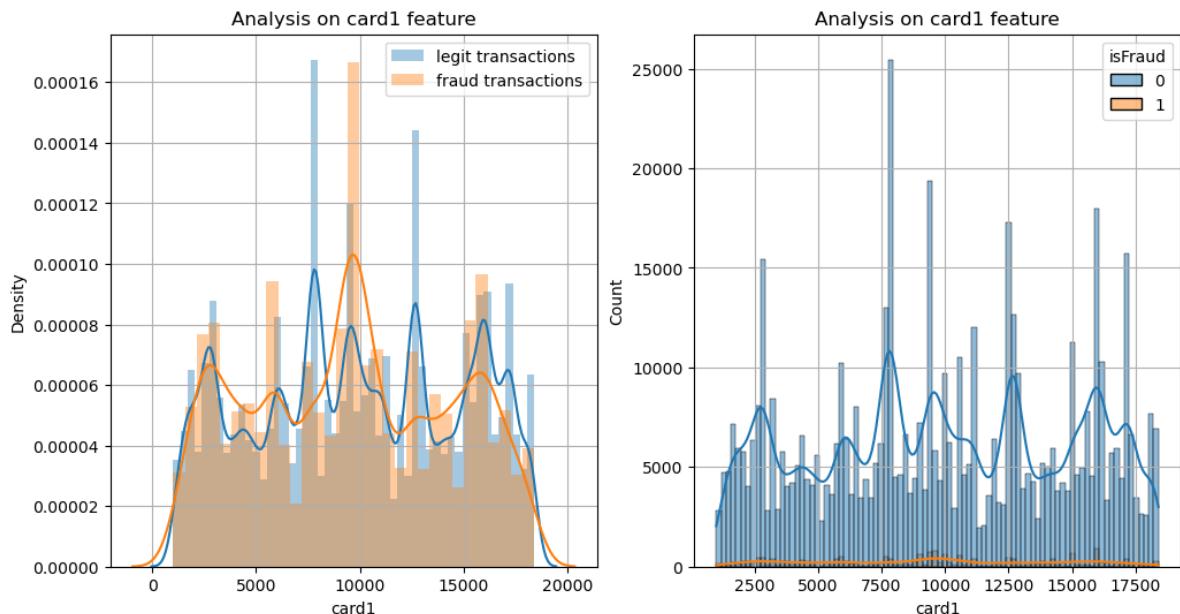
Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	590540.000000	count	569877.000000	
mean	9898.734658	mean	9911.464381	
std	4901.170153	std	4905.672370	
min	1000.000000	min	1000.000000	
25%	6019.000000	25%	6019.000000	
50%	5732.500000			
75%	9633.000000	50%	9711.000000	
max	14184.000000	75%	14229.000000	
	13623.000000	max	18396.000000	
	18375.000000			
Name: card1, dtype: float64		Name: card1, dtype: float64		
Name: card1, dtype: float64				

The unique values in card1 feature is: 13553
 card1 feature has 0 null values, which holds 0.0 %



	card1	isFraud	total	% Occurence	Avg Fraud
5365	7919	112	14932	0.025285	0.007501
6615	9500	528	14162	0.023981	0.037283
11593	15885	444	10361	0.017545	0.042853
12616	17188	278	10344	0.017516	0.026875
10950	15066	313	7945	0.013454	0.039396
...
7157	10200	0	1	0.000002	0.000000
7151	10191	0	1	0.000002	0.000000
7148	10187	0	1	0.000002	0.000000
7145	10184	0	1	0.000002	0.000000
13552	18396	0	1	0.000002	0.000000

[13553 rows x 5 columns]



```
In [22]: # to get a dataframe with respect to corresponding feature and target variable
temp = pd.DataFrame(df_train.groupby('card1')[['isFraud']].agg(lambda x: x.eq(1).sum))

temp['total'] = pd.DataFrame(df_train.groupby('card1')[['isFraud']].agg(total='count'))
temp['% Occurence'] = temp['total'] / df_train.shape[0]
temp['Avg Fraud'] = pd.DataFrame(df_train.groupby('card1')[['isFraud']].agg(Avg='mean'))

temp.sort_values(by=[ 'Avg Fraud'], inplace=True, ascending=False)
```

```
In [23]: # to find out card transactions in which fraudulent rate is more than 70% and total
cond = temp[temp['Avg Fraud'] >= 0.70]
print(cond[cond['total'] >= 3])
```

1. self_case_study_1_IEEE_fraud_detection_EDA

	card1	isFraud	total	% Occurrence	Avg Fraud
7259	10320	4	4	0.000007	1.000000
10655	14671	3	3	0.000005	1.000000
10930	15042	5	5	0.000008	1.000000
2920	4774	6	6	0.000010	1.000000
11099	15253	7	7	0.000012	1.000000
9452	13103	3	3	0.000005	1.000000
3166	5093	3	3	0.000005	1.000000
2480	4179	5	5	0.000008	1.000000
7822	11031	6	6	0.000010	1.000000
1168	2497	4	4	0.000007	1.000000
2673	4448	5	5	0.000008	1.000000
6539	9401	3	3	0.000005	1.000000
4832	7228	3	3	0.000005	1.000000
8151	11450	3	3	0.000005	1.000000
858	2089	4	4	0.000007	1.000000
5702	8341	3	3	0.000005	1.000000
9810	13549	4	4	0.000007	1.000000
5851	8525	3	3	0.000005	1.000000
3820	5942	4	4	0.000007	1.000000
3375	5350	3	3	0.000005	1.000000
326	1394	7	7	0.000012	1.000000
10274	14157	4	4	0.000007	1.000000
8292	11625	3	3	0.000005	1.000000
1833	3342	10	10	0.000017	1.000000
7079	10101	3	3	0.000005	1.000000
11744	16083	3	3	0.000005	1.000000
1239	2591	3	3	0.000005	1.000000
13046	17739	8	8	0.000014	1.000000
8957	12473	46	47	0.000080	0.978723
8374	11716	16	17	0.000029	0.941176
1308	2675	9	10	0.000017	0.900000
2400	4078	7	8	0.000014	0.875000
154	1191	6	7	0.000012	0.857143
1655	3123	6	7	0.000012	0.857143
7476	10586	6	7	0.000012	0.857143
3020	4904	5	6	0.000010	0.833333
3861	5993	5	6	0.000010	0.833333
6673	9579	5	6	0.000010	0.833333
8049	11331	9	11	0.000019	0.818182
7389	10487	4	5	0.000008	0.800000
383	1469	8	10	0.000017	0.800000
2115	3710	4	5	0.000008	0.800000
5038	7492	4	5	0.000008	0.800000
12352	16859	7	9	0.000015	0.777778
3754	5853	23	30	0.000051	0.766667
12309	16801	3	4	0.000007	0.750000
10436	14364	3	4	0.000007	0.750000
8716	12160	3	4	0.000007	0.750000
6916	9890	3	4	0.000007	0.750000
4090	6280	3	4	0.000007	0.750000
2117	3712	3	4	0.000007	0.750000
527	1657	3	4	0.000007	0.750000
8916	12422	9	12	0.000020	0.750000
1172	2502	3	4	0.000007	0.750000
6072	8801	11	15	0.000025	0.733333
542	1676	8	11	0.000019	0.727273
8609	12023	13	18	0.000030	0.722222

```
In [24]: nfr = len(cond[cond['total'] >=3])
print('Number of cases where transaction been made more than thrice and chances of that b
```

eing fraud are more than 70%: 57

```
In [25]: print('Number of card transactions when the likelihood of fraud is 100 percent:',  
      print('Number of card transactions where there are more than 70% fraud cases:', len(  
      print(f'There are {nfr} such cases where card transaction occurred three or more than  
      three times and the chances of that being fraud are more than 70%'))
```

Number of card transactions when the likelihood of fraud is 100 percent: 203
 Number of card transactions where there are more than 70% fraud cases: 232
 There are 57 such cases where card transaction occurred three or more than three times and the chances of that being fraud are more than 70%

Observation: We know that feature 'card1' is a numerical categorical variable, though the values are integer but they don't hold any ordinal significance, they are like categories. We deduced from the feature description that the distribution of legitimate transactions and fraudulent transactions is nearly same. The distribution plot showed that there are more fraudulent transactions involving cards with numbers between 9000 and 10000.

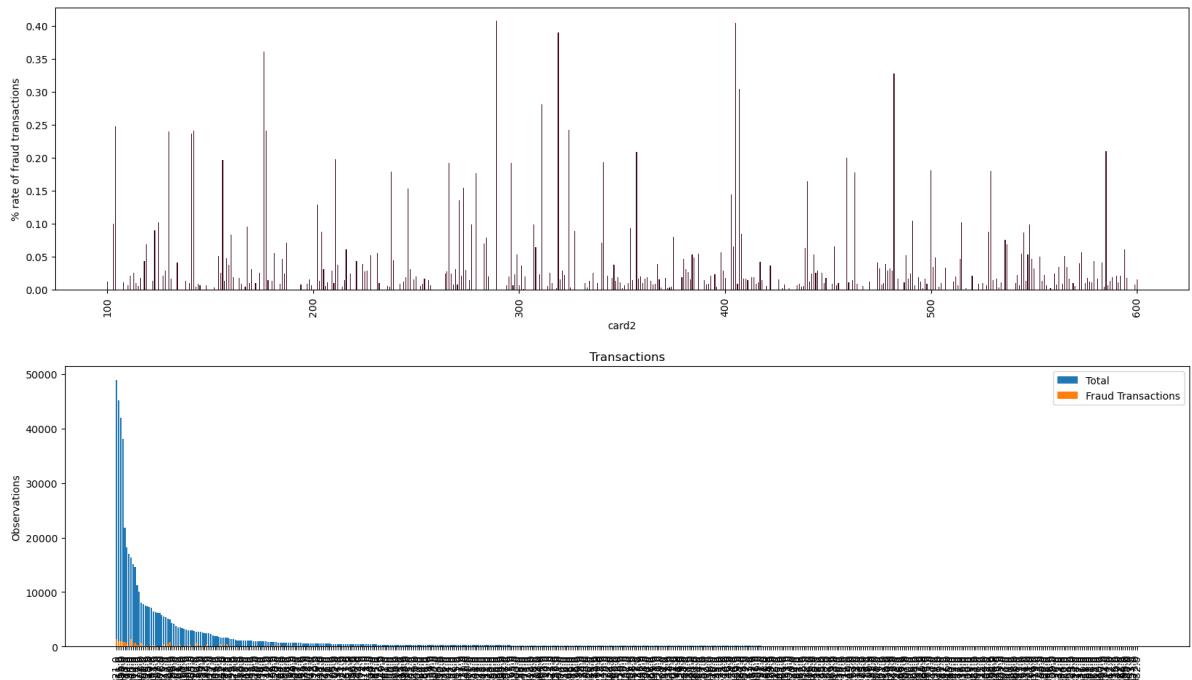
1.1.3.2 Card2

```
In [26]: plot_categorical_features(df_train, 'card2', 'isFraud')
```

Five Number Summary			
Feature Description of legit transactions		Feature Description of fraud transactions	
count	581607.000000	count	561367.000000
mean	362.555488	mean	362.453981
std	157.793246	std	157.726065
min	100.000000	min	100.000000
25%	214.000000	25%	215.000000
50%	361.000000	50%	361.000000
75%	512.000000	75%	512.000000
max	600.000000	max	600.000000
	Name: card2, dtype: float64		Name: card2, dtype: float64
	Name: card2, dtype: float64		

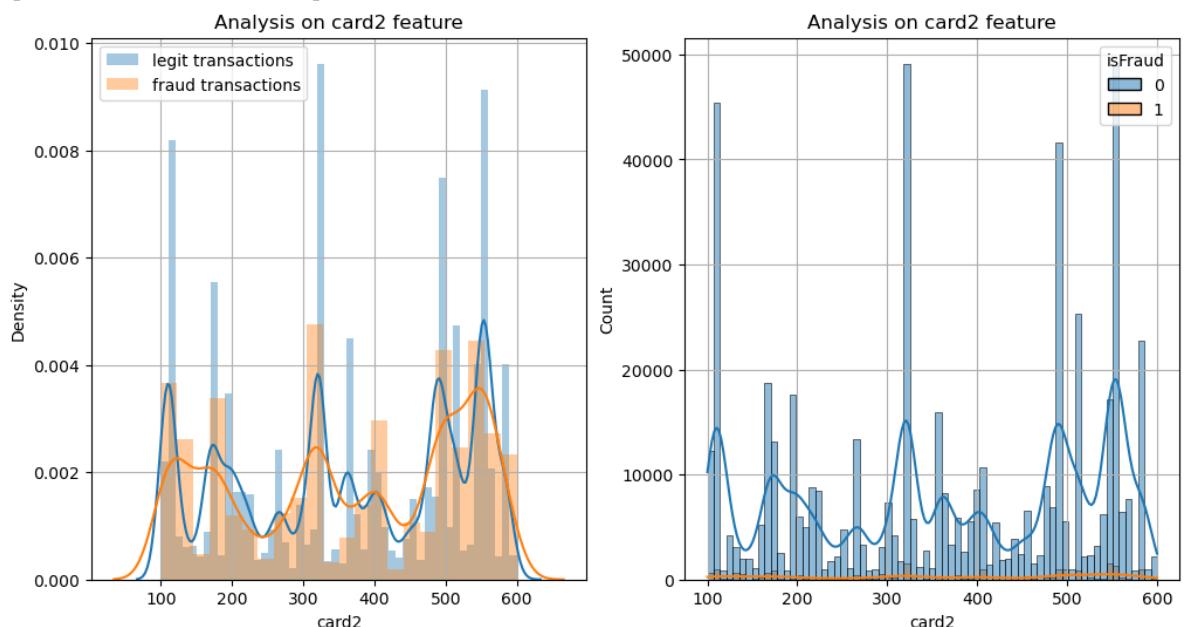
The unique values in card2 feature is: 500
 card2 feature has 8933 null values, which holds 0.015126833068039422 %

1. self_case_study_1_IEEE_fraud_detection_EDA



	card2	isFraud	total	% Occurrence	Avg Fraud
220	321.0	1396	48935	0.082865	0.028528
10	111.0	978	45191	0.076525	0.021641
454	555.0	959	41995	0.071113	0.022836
389	490.0	916	38145	0.064593	0.024014
482	583.0	899	21803	0.036920	0.041233
..
346	447.0	1	39	0.000066	0.025641
52	153.0	0	38	0.000064	0.000000
456	557.0	0	37	0.000063	0.000000
372	473.0	0	18	0.000030	0.000000
481	582.0	0	14	0.000024	0.000000

[500 rows x 5 columns]



```
In [27]: # to get a dataframe with respect to corresponding feature and target variable
temp = pd.DataFrame(df_train.groupby('card2')['isFraud'].agg(lambda x: x.eq(1).sum))

temp['total'] = pd.DataFrame(df_train.groupby('card2')['isFraud'].agg(total='count'))
temp['% Occurrence'] = temp['total'] / df_train.shape[0]
temp['Avg Fraud'] = pd.DataFrame(df_train.groupby('card2')['isFraud'].agg(Avg='mean'))
```

```
temp.sort_values(by=['Avg Fraud'], inplace=True, ascending=False)
```

In [28]: temp

	card2	isFraud	total	% Occurrence	Avg Fraud
188	289.0	48	118	0.000200	0.406780
304	405.0	92	228	0.000386	0.403509
218	319.0	39	100	0.000169	0.390000
75	176.0	185	513	0.000869	0.360624
381	482.0	55	168	0.000284	0.327381
...
385	486.0	0	121	0.000205	0.000000
285	386.0	0	83	0.000141	0.000000
88	189.0	0	106	0.000179	0.000000
87	188.0	0	80	0.000135	0.000000
100	201.0	0	164	0.000278	0.000000

500 rows × 5 columns

Observation: Though the numbers are integers, we know that feature "card2" is a numerical categorical variable; as such, even though the values are integers, they do not have any ordinal meaning. From the feature description, we inferred that the distribution of legit and fraudulent transactions is about equal. In contrast to the "card1" feature, a transaction can be fraudulent up to 40% of the time here. No transactions (with card 2 number) were reported as having a fraud rate of greater than 40%.

1.1.3.3 Card3

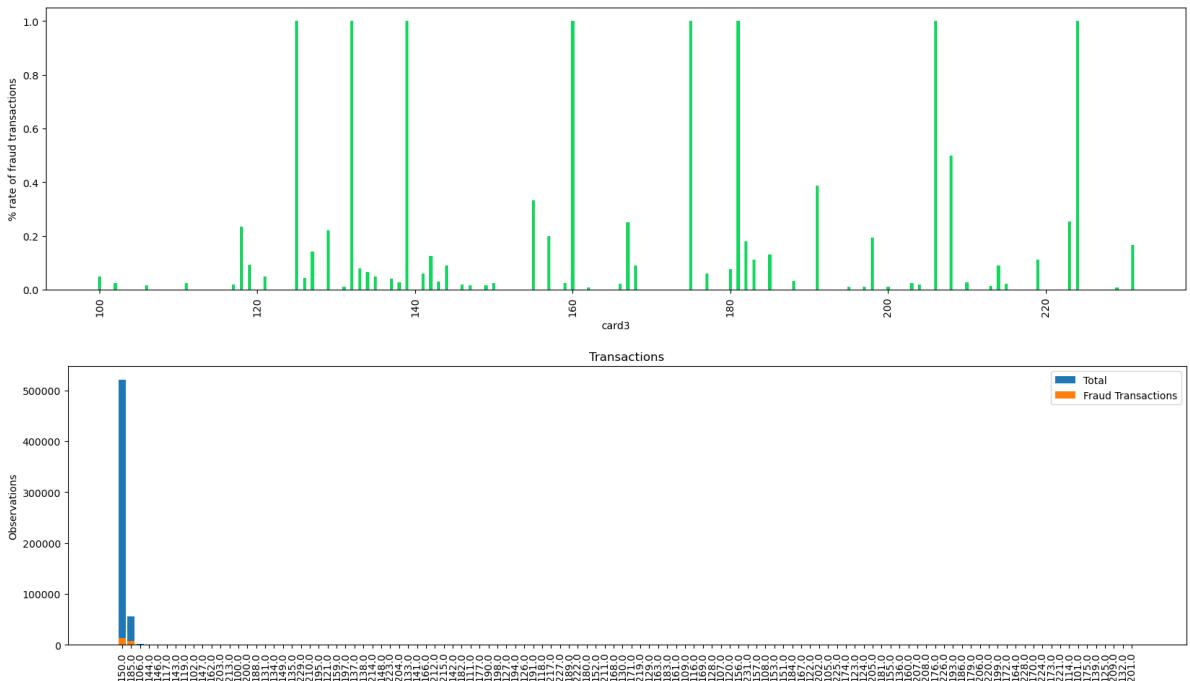
In [29]: plot_categorical_features(df_train, 'card3', 'isFraud')

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	588975.000000	count	568351.000000	
mean	153.194925	mean	152.862034	
std	11.336444	std	10.904432	
min	100.000000	min	100.000000	
25%	150.000000	25%	150.000000	
50%	150.000000	50%	150.000000	
75%	150.000000	75%	150.000000	
max	231.000000	max	231.000000	
Name: card3, dtype: float64		Name: card3, dtype: float64		
Name: card3, dtype: float64				

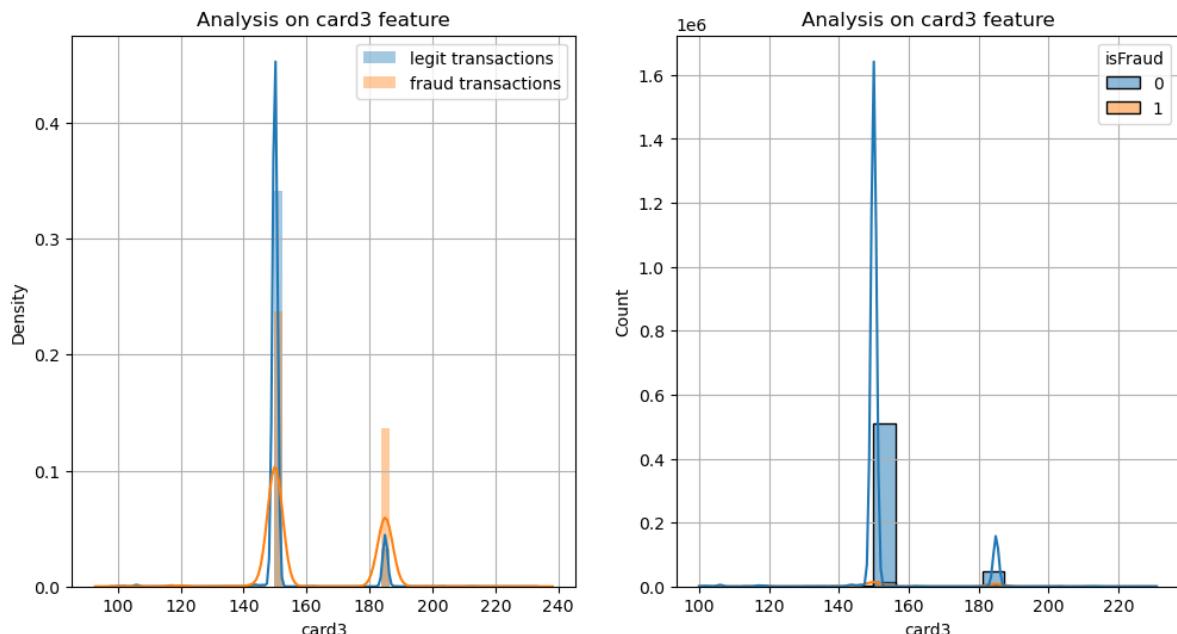
The unique values in card3 feature is: 114

card3 feature has 1565 null values, which holds 0.0026501168422122124 %



	card3	isFraud	total	% Occurence	Avg Fraud
42	150.0	12807	521287	0.882729	0.024568
73	185.0	7362	56346	0.095414	0.130657
4	106.0	24	1571	0.002660	0.015277
37	144.0	112	1252	0.002120	0.089457
38	146.0	25	1252	0.002120	0.019968
..
33	139.0	1	1	0.000002	1.000000
19	125.0	1	1	0.000002	1.000000
94	209.0	0	1	0.000002	0.000000
26	132.0	1	1	0.000002	1.000000
86	201.0	0	1	0.000002	0.000000

[114 rows x 5 columns]



```
In [30]: # to get a dataframe with respect to corresponding feature and target variable
temp = pd.DataFrame(df_train.groupby('card3')[['isFraud']].agg(lambda x: x.eq(1).sum))

temp['total'] = pd.DataFrame(df_train.groupby('card3')[['isFraud']].agg(total='count'))
temp['% Occurence'] = temp['total'] / df_train.shape[0]
temp['Avg Fraud'] = pd.DataFrame(df_train.groupby('card3')[['isFraud']].agg(Avg='mean'))

temp.sort_values(by=['Avg Fraud'], inplace=True, ascending=False)
temp
```

Out[30]:

	card3	isFraud	total	% Occurrence	Avg Fraud
33	139.0	1	1	0.000002	1.0
50	160.0	3	3	0.000005	1.0
107	224.0	1	1	0.000002	1.0
64	175.0	1	1	0.000002	1.0
69	181.0	3	3	0.000005	1.0
...
74	186.0	0	2	0.000003	0.0
24	130.0	0	10	0.000017	0.0
76	189.0	0	14	0.000024	0.0
9	114.0	0	1	0.000002	0.0
14	120.0	0	6	0.000010	0.0

114 rows × 5 columns

In [31]: `print('card transactions where there are more than 100% fraud cases:', len(temp[temp['isFraud'] > 1].groupby('card3').size() / temp['isFraud'].sum() * 100 > 100))`

card transactions where there are more than 100% fraud cases: 8

Observation: There are only 115 categories in "card3," with 150.0 occurring the most frequently (521287) and followed by 185.0 . According to the distribution plot, the majority of the transactions that are flagged as fraudulent fall between the range of 150.0–155.0 and 180.0–190.0 .

1.1.3.4 Card4

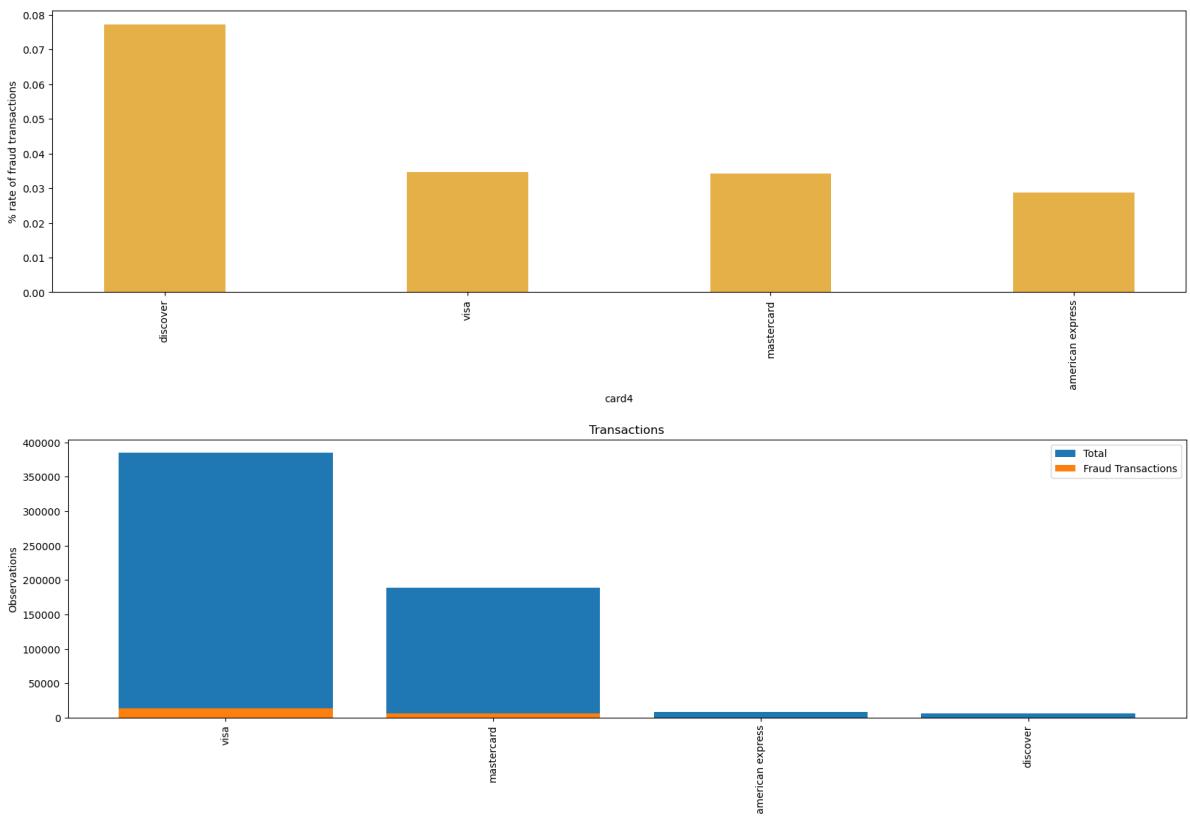
In [32]: `plot_categorical_features(df_train, 'card4', 'isFraud')`

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	588963	count	568341	
unique	20622	unique	4	
top	visa	top	visa	
freq	384767	freq	371394	
Name: card4, dtype: object		Name: card4, dtype: object		
Name: card4, dtype: object				

The unique values in card4 feature is: 4

card4 feature has 1577 null values, which holds 0.00267043722694483 %



	card4	isFraud	total	% Occurrence	Avg Fraud
3	visa	13373	384767	0.651551	0.034756
2	mastercard		6496	0.320414	0.034331
0	american express		239	0.014102	0.028698
1	discover		514	0.011263	0.077282

Observation: The feature "card4" has a total of 4 categories. Visa is the category that pops up here the most frequently. It happens at 65.15 %. The least frequent category is "discover," yet there are more instances of fraudulent transactions in this group. If the transaction falls into the "discover" category, there is a 7.73% possibility that it is fraudulent.

1.1.3.5 Card5

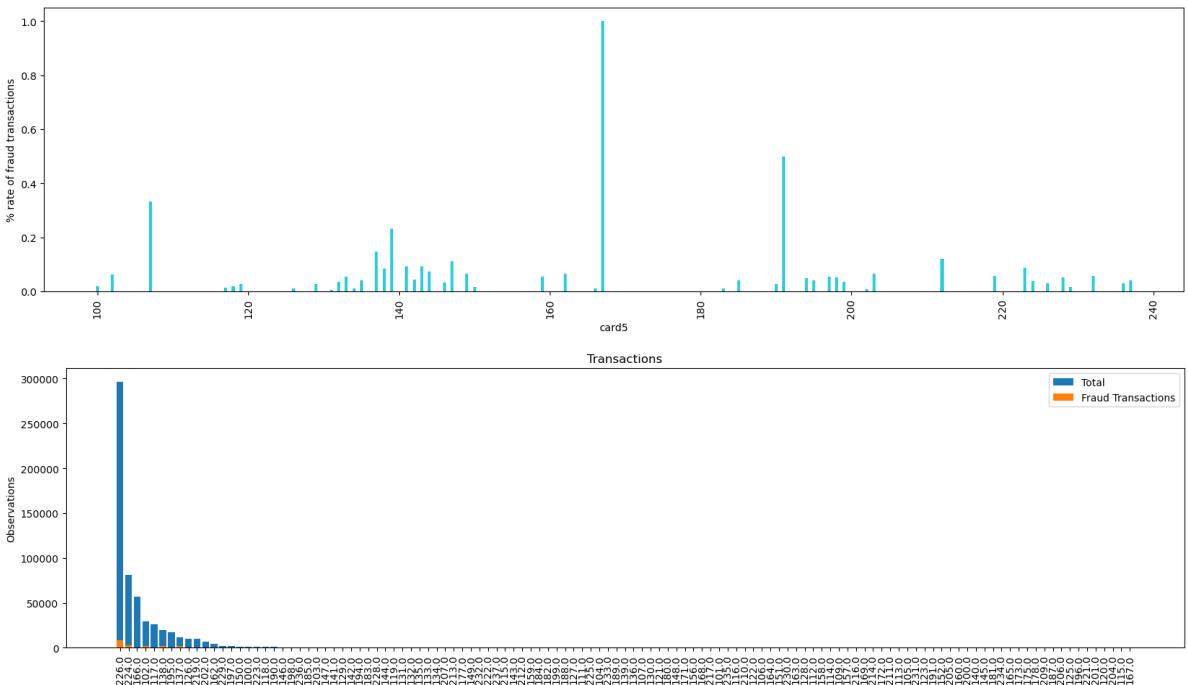
```
In [33]: plot_categorical_features(df_train, 'card5', 'isFraud')
```

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
		count	mean	count
count	586281.000000	20453.000000	199.278897	565828.000000
mean	191.994328	199.542212	199.542212	199.542212
std	41.244453	41.046975	41.046975	41.046975
std	45.779886			
min	100.000000	100.000000	100.000000	100.000000
min	100.000000			
25%	166.000000	166.000000	166.000000	166.000000
25%	138.000000			
50%	224.000000	226.000000	226.000000	226.000000
50%	224.000000			
75%	226.000000	226.000000	226.000000	226.000000
75%	226.000000			
max	237.000000	237.000000	237.000000	237.000000
max	237.000000			
Name: card5, dtype: float64		Name: card5, dtype: float64		
Name: card5, dtype: float64				

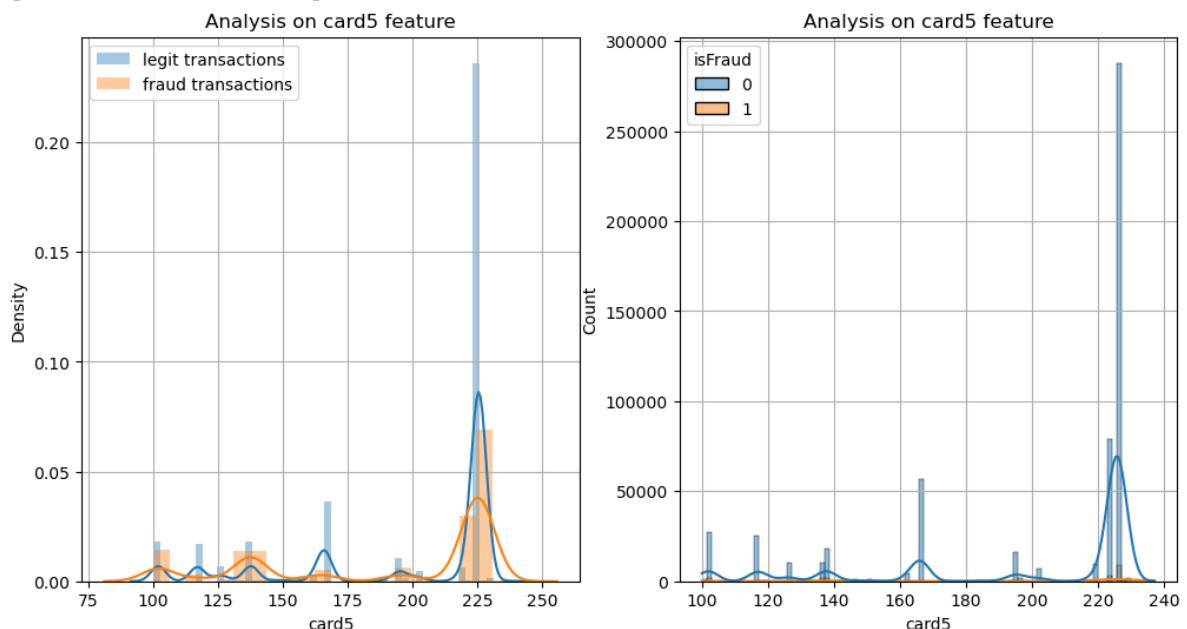
The unique values in card5 feature is: 119

card5 feature has 4259 null values, which holds 0.007212043214684865 %



	card5	isFraud	total	% Occurence	Avg Fraud
108	226.0	8753	296546	0.502161	0.029516
106	224.0	3152	81513	0.138031	0.038669
58	166.0	630	57140	0.096759	0.011026
2	102.0	1832	29105	0.049285	0.062945
14	117.0	354	25941	0.043928	0.013646
..
86	201.0	0	1	0.000002	0.000000
17	120.0	0	1	0.000002	0.000000
89	204.0	0	1	0.000002	0.000000
12	115.0	0	1	0.000002	0.000000
59	167.0	1	1	0.000002	1.000000

[119 rows x 5 columns]



Observation: In "card5", there are only 120 categories, and card no. 226.0 appear more than 50% of the time. The majority of the transactions that are marked as fraudulent fall between the ranges of 215.0-235.0, 100.0-110.0, and 130.0-145.0, as per the distribution plot.

1.1.3.6 Card6

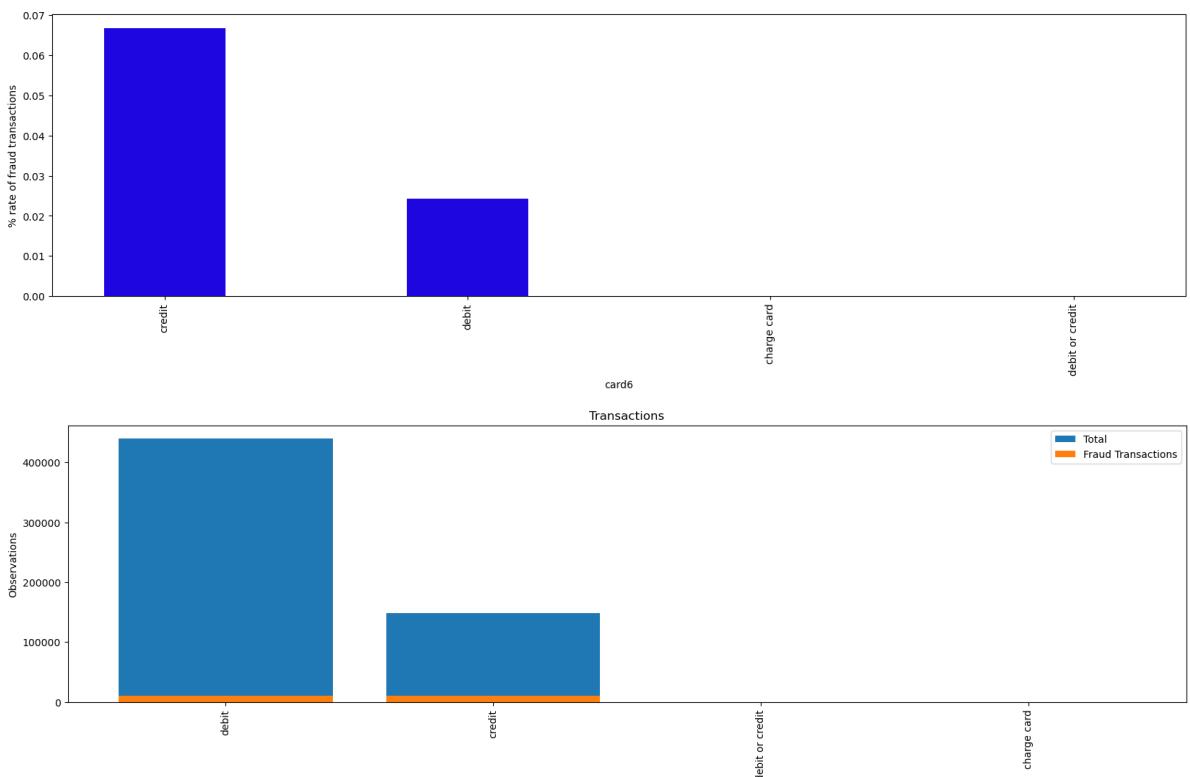
```
In [34]: plot_categorical_features(df_train, 'card6', 'isFraud')
```

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	588969	count	568345	
unique	20624	unique	4	
top	debit	top	debit	
freq	439938	freq	429264	
	10674			Name: card6, dtype: object
				Name: card6, dtype: object

The unique values in card6 feature is: 4

card6 feature has 1571 null values, which holds 0.0026602770345785214 %



	card6	isFraud	total	% Occurence	Avg Fraud
2	debit	10674	439938	0.744976	0.024263
1	credit	9950	148986	0.252288	0.066785
3	debit or credit	0	30	0.000051	0.000000
0	charge card	0	15	0.000025	0.000000

Observation: The feature "card6" has a total of 4 unique categories excluding the null values out of which 'debit or credit' and 'charge card' occurs very few times, almost negligible times. From the observation it has been clear that credit card transactions are being marked as fraudulent more than debit card transactions.

1.1.4 Address Features

1.1.4.1 addr1

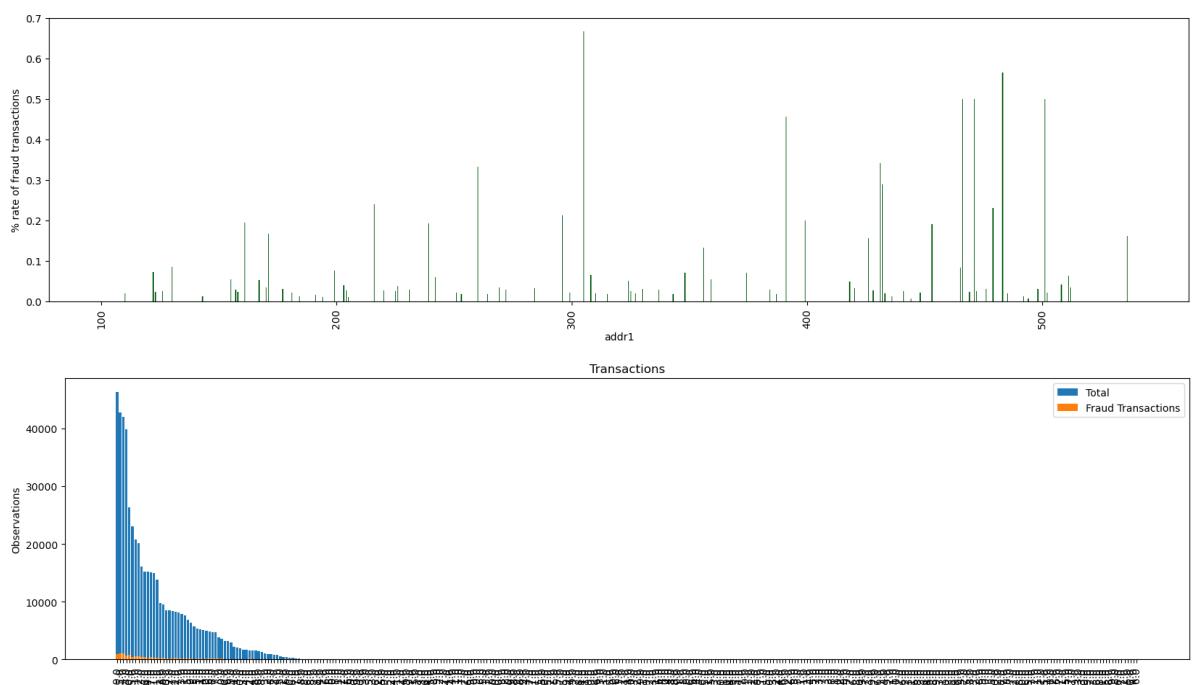
```
In [35]: plot_categorical_features(df_train, 'addr1', 'isFraud')
```

Five Number Summary

	Feature Description e Description of fraud transactions	Feature Description of legit transactions	Featur
count	524834.000000	count	511912.000000
mean	290.733794	mean	290.643341
std	294.317134	std	101.692564
min	101.741072	min	103.584955
25%	100.000000	25%	100.000000
50%	110.000000	50%	100.000000
75%	204.000000	75%	204.000000
max	299.000000	max	299.000000
	Name: addr1, dtype: float64	Name: addr1, dtype: float64	
	Name: addr1, dtype: float64		

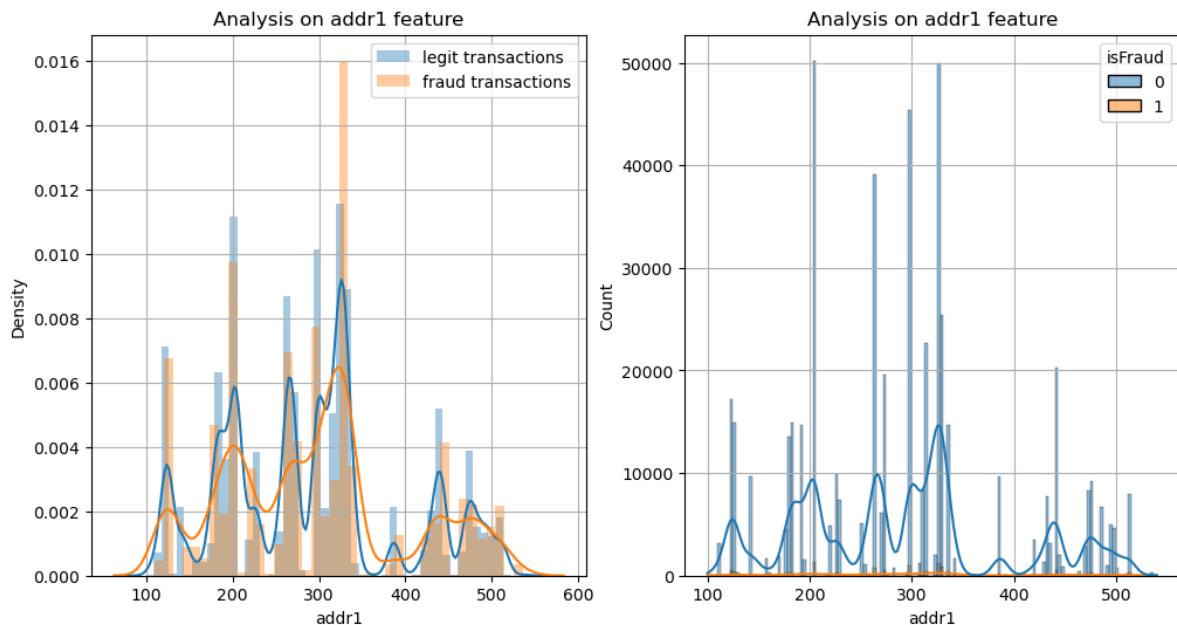
The unique values in addr1 feature is: 332

addr1 feature has 65706 null values, which holds 0.1112642666034477 %



	addr1	isFraud	total	% Occurrence	Avg Fraud
151	299.0	985	46335	0.078462	0.021258
173	325.0	1087	42751	0.072393	0.025426
78	204.0	1120	42020	0.071155	0.026654
127	264.0	728	39870	0.067514	0.018259
178	330.0	840	26287	0.044513	0.031955
..
215	382.0	0	1	0.000002	0.000000
111	245.0	0	1	0.000002	0.000000
218	386.0	0	1	0.000002	0.000000
110	244.0	0	1	0.000002	0.000000
80	206.0	0	1	0.000002	0.000000

[332 rows x 5 columns]



```
In [36]: # to get a dataframe with respect to corresponding feature and target variable
temp = pd.DataFrame(df_train.groupby('addr1')['isFraud'].agg(lambda x: x.eq(1).sum))

temp['total'] = pd.DataFrame(df_train.groupby('addr1')['isFraud'].agg(total='count'))
temp['% Occurrence'] = temp['total'] / df_train.shape[0]
temp['Avg Fraud'] = pd.DataFrame(df_train.groupby('addr1')['isFraud'].agg(Avg='mean'))

temp.sort_values(by=['Avg Fraud'], inplace=True, ascending=False)
temp
```

Out[36]:

	addr1	isFraud	total	% Occurrence	Avg Fraud
157	305.0	6	9	0.000015	0.666667
288	483.0	13	23	0.000039	0.565217
301	501.0	2	4	0.000007	0.500000
274	466.0	1	2	0.000003	0.500000
279	471.0	1	2	0.000003	0.500000
...
136	277.0	0	3	0.000005	0.000000
137	278.0	0	1	0.000002	0.000000
138	279.0	0	1	0.000002	0.000000
139	280.0	0	1	0.000002	0.000000
331	540.0	0	6	0.000010	0.000000

332 rows × 5 columns

Observation: There were just two instances that had a chance of a fraud transaction above 50%. Since the distribution of legitimate and fraudulent transactions virtually overlaps for this feature, it is difficult to draw firm conclusions.

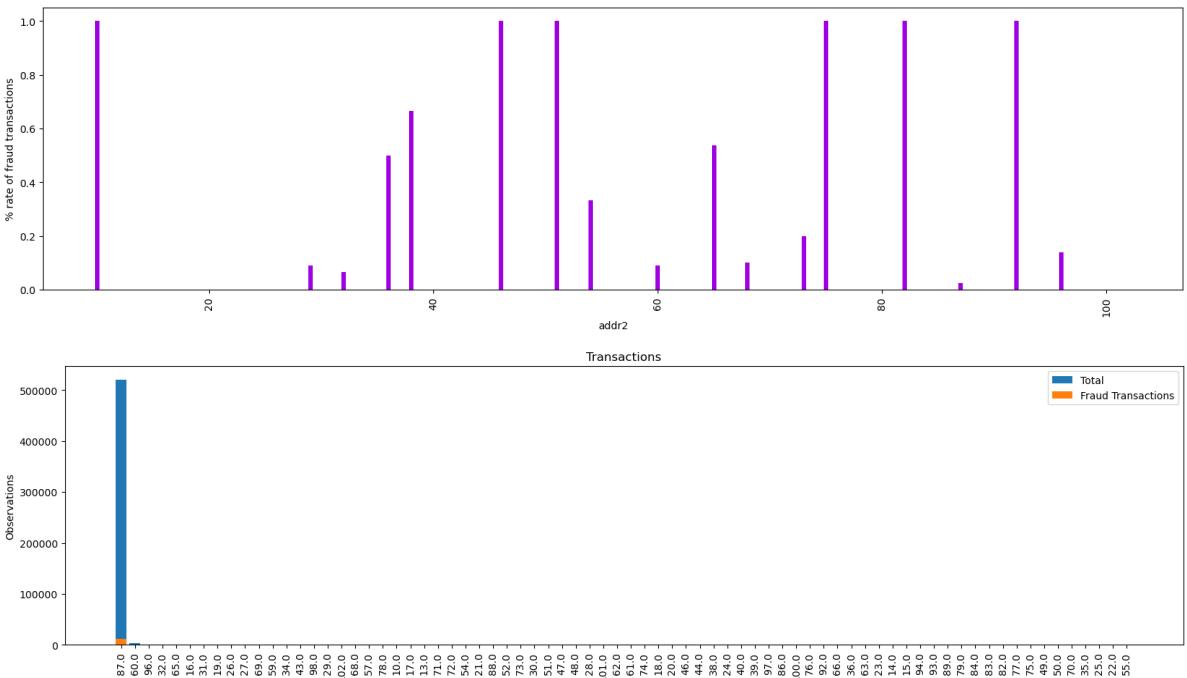
1.1.4.2 addr2

In [37]: `plot_categorical_features(df_train, 'addr2', 'isFraud')`

Five Number Summary

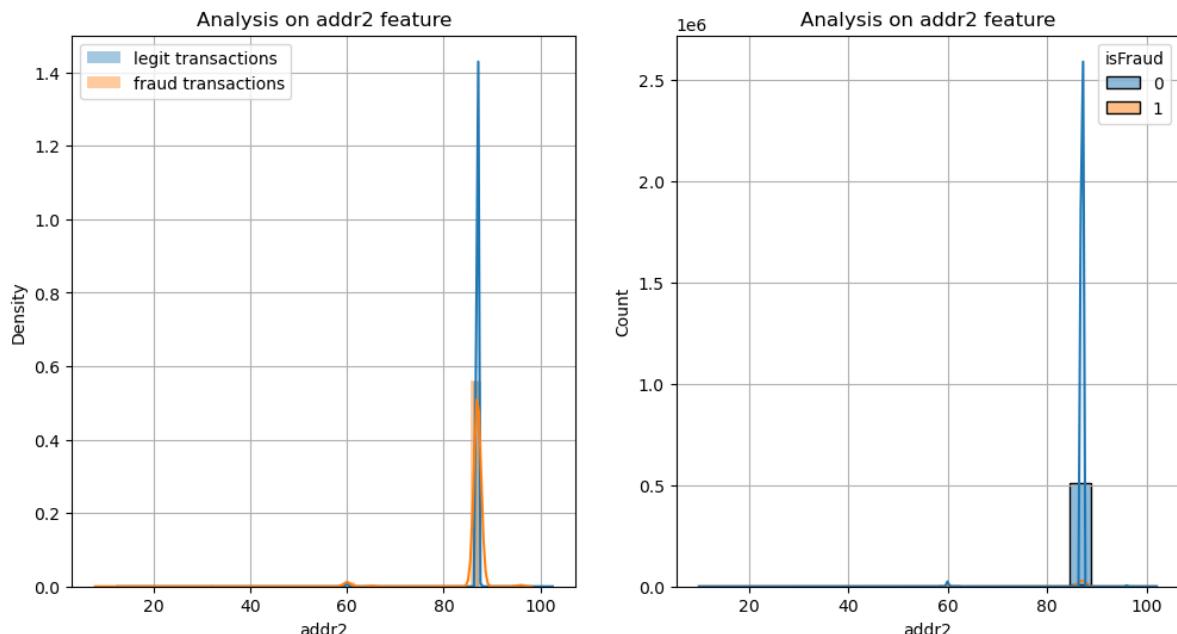
Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count 524834.000000		count 511912.000000		count 12922.000000
mean 86.800630		mean 86.813620		mean 86.286024
std 2.690623		std 2.607742		std 4.936238
min 10.000000		min 13.000000		min 10.000000
25% 87.000000		25% 87.000000		25% 87.000000
50% 87.000000		50% 87.000000		50% 87.000000
75% 87.000000		75% 87.000000		75% 87.000000
max 102.000000		max 102.000000		max 96.000000
Name: addr2, dtype: float64		Name: addr2, dtype: float64		Name: addr2, dtype: float64
Name: addr2, dtype: float64				

The unique values in addr2 feature is: 74
addr2 feature has 65706 null values, which holds 0.1112642666034477 %



	addr2	isFraud	total	% Occurence	Avg Fraud
62	87.0	12477	520481	0.881365	0.023972
40	60.0	279	3084	0.005222	0.090467
68	96.0	89	638	0.001080	0.139498
20	32.0	6	91	0.000154	0.065934
44	65.0	44	82	0.000139	0.536585
..
48	70.0	0	1	0.000002	0.000000
22	35.0	0	1	0.000002	0.000000
13	25.0	0	1	0.000002	0.000000
10	22.0	0	1	0.000002	0.000000
37	55.0	0	1	0.000002	0.000000

[74 rows x 5 columns]



```
In [38]: # to get a dataframe with respect to corresponding feature and target variable
temp = pd.DataFrame(df_train.groupby('addr2')['isFraud'].agg(lambda x: x.eq(1).sum))

temp['total'] = pd.DataFrame(df_train.groupby('addr2')['isFraud'].agg(total='count'))
temp['% Occurence'] = temp['total'] / df_train.shape[0]
temp['Avg Fraud'] = pd.DataFrame(df_train.groupby('addr2')['isFraud'].agg(Avg='mean'))

temp.sort_values(by=['Avg Fraud'], inplace=True, ascending=False)
temp
```

Out[38]:

	addr2	isFraud	total	% Occurrence	Avg Fraud
0	10.0	8	8	0.000014	1.0
58	82.0	1	1	0.000002	1.0
34	51.0	4	4	0.000007	1.0
29	46.0	3	3	0.000005	1.0
65	92.0	2	2	0.000003	1.0
...
32	49.0	0	1	0.000002	0.0
33	50.0	0	1	0.000002	0.0
35	52.0	0	5	0.000008	0.0
1	13.0	0	7	0.000012	0.0
73	102.0	0	11	0.000019	0.0

74 rows × 5 columns

In [39]: `print('Number of card transactions when the likelihood of fraud is 100 percent: ',`

Number of card transactions when the likelihood of fraud is 100 percent: 6

In [40]: `temp`

Out[40]:

	addr2	isFraud	total	% Occurrence	Avg Fraud
0	10.0	8	8	0.000014	1.0
58	82.0	1	1	0.000002	1.0
34	51.0	4	4	0.000007	1.0
29	46.0	3	3	0.000005	1.0
65	92.0	2	2	0.000003	1.0
...
32	49.0	0	1	0.000002	0.0
33	50.0	0	1	0.000002	0.0
35	52.0	0	5	0.000008	0.0
1	13.0	0	7	0.000012	0.0
73	102.0	0	11	0.000019	0.0

74 rows × 5 columns

Observation: Category 87.0 occurs almost 88.13% of the time, showing that this feature is heavily predominated by this category. From the distribution, we deduced that the distribution of fraudulent transactions from the 75th to the 100th percentile slightly differs from the distribution of legit transactions. The maximum value of a fraudulent transaction is 96.0, while the maximum value of a legitimate transaction is 102.0. i.e, if a "addr2" value is more than 96.00, there is a 100% likelihood that it is a legit transaction.

1.1.5 Email Features

1.1.5.1 P_emaildomain

```
In [41]: cat_fea = ['ProductCD', 'card1', 'card2', 'card3', 'card4', 'card5', 'card6', 'addr1', 'P_emaildomain', 'R_emaildomain', 'M1', 'M2', 'M3', 'M4', 'M5', 'M6', 'I1', 'DeviceType', 'DeviceInfo', 'id_12', 'id_13', 'id_14', 'id_15', 'id_16', 'id_19', 'id_20', 'id_21', 'id_22', 'id_23', 'id_24', 'id_25', 'id_26', 'id_29', 'id_30', 'id_31', 'id_32', 'id_33', 'id_34', 'id_35', 'id_36']
```

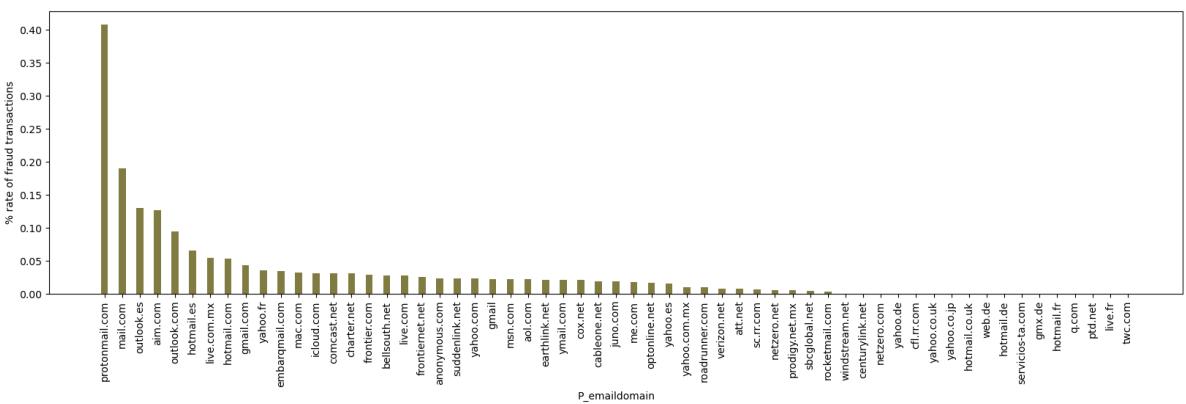
```
In [42]: plot_categorical_features(df_train, 'P_emaildomain', 'isFraud')
```

Five Number Summary

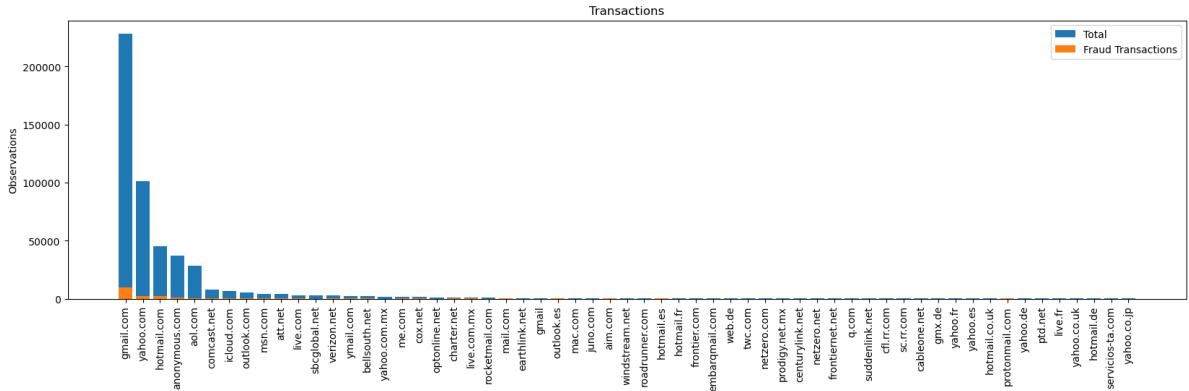
	Feature Description		Feature Description of legit transactions	
	Feature Description of fraud transactions			
count	count	496084	count	478211
unique	unique	59	unique	59
freq	top	gmail.com	top	gmail.com
	top	gmail.com		
	freq	228355	freq	218412
	Name: P_emaildomain, dtype: object		Name: P_emaildomain, dtype: object	
	Name: P_emaildomain, dtype: object			

The unique values in P_emaildomain feature is: 59

P_emaildomain feature has 94456 null values, which holds 0.1599485216920107 %



1. self_case_study_1_IEEE_fraud_detection_EDA



	P_emaildomain	isFraud	total	% Occurence	Avg Fraud
16	gmail.com	9943	228355	0.386688	0.043542
53	yahoo.com	2297	100934	0.170918	0.022757
19	hotmail.com	2396	45250	0.076625	0.052950
1	anonymous.com	859	36998	0.062651	0.023217
2	aol.com	617	28289	0.047904	0.021811
9	comcast.net	246	7888	0.013357	0.031187
23	icloud.com	197	6267	0.010612	0.031434
35	outlook.com	482	5096	0.008629	0.094584
31	msn.com	90	4092	0.006929	0.021994
3	att.net	30	4033	0.006829	0.007439
25	live.com	84	3041	0.005150	0.027622
43	sbcglobal.net	12	2970	0.005029	0.004040
48	verizon.net	22	2705	0.004581	0.008133
58	ymail.com	50	2396	0.004057	0.020868
4	bellsouth.net	53	1909	0.003233	0.027763
54	yahoo.com.mx	16	1543	0.002613	0.010369
30	me.com	27	1522	0.002577	0.017740
10	cox.net	29	1393	0.002359	0.020818
34	optonline.net	17	1011	0.001712	0.016815
8	charter.net	25	816	0.001382	0.030637
26	live.com.mx	41	749	0.001268	0.054740
42	rocketmail.com	2	664	0.001124	0.003012
29	mail.com	106	559	0.000947	0.189624
11	earthlink.net	11	514	0.000870	0.021401
15	gmail	11	496	0.000840	0.022177
36	outlook.es	57	438	0.000742	0.130137
28	mac.com	14	436	0.000738	0.032110
24	juno.com	6	322	0.000545	0.018634
0	aim.com	40	315	0.000533	0.126984
50	windstream.net	0	305	0.000516	0.000000
41	roadrunner.com	3	305	0.000516	0.009836
21	hotmail.es	20	305	0.000516	0.065574
22	hotmail.fr	0	295	0.000500	0.000000
13	frontier.com	8	280	0.000474	0.028571
12	embarqmail.com	9	260	0.000440	0.034615
49	web.de	0	240	0.000406	0.000000
47	twc.com	0	230	0.000389	0.000000
32	netzero.com	0	230	0.000389	0.000000
37	prodigy.net.mx	1	207	0.000351	0.004831
6	centurylink.net	0	205	0.000347	0.000000
33	netzero.net	1	196	0.000332	0.005102
14	frontiernet.net	5	195	0.000330	0.025641
40	q.com	0	189	0.000320	0.000000
46	suddenlink.net	4	175	0.000296	0.022857
7	cfl.rr.com	0	172	0.000291	0.000000
44	sc.rr.com	1	164	0.000278	0.006098
5	cableone.net	3	159	0.000269	0.018868
17	gmx.de	0	149	0.000252	0.000000
57	yahoo.fr	5	143	0.000242	0.034965
56	yahoo.es	2	134	0.000227	0.014925
18	hotmail.co.uk	0	112	0.000190	0.000000
38	protonmail.com	31	76	0.000129	0.407895
55	yahoo.de	0	74	0.000125	0.000000
39	ptd.net	0	68	0.000115	0.000000
27	live.fr	0	56	0.000095	0.000000
52	yahoo.co.uk	0	49	0.000083	0.000000
20	hotmail.de	0	43	0.000073	0.000000
45	servicios-ta.com	0	35	0.000059	0.000000
51	yahoo.co.jp	0	32	0.000054	0.000000

Observation: This feature has 59 categories altogether. Naturally, "gmail.com" is the category that pops up the most in this, occurring 38.67% of the time, followed by "yahoo.com". We observed from the barplot that there is a greater than 40% chance that a transaction made through "protonmail.com" is fraudulent. after that came "mail.com" (18%) and "outlook.es" (13%). Additionally, we noticed that the distribution of legitimate transactions falls into 59 distinct categories, whereas the distribution of fraudulent transactions falls into 42 distinct categories, leading us to the conclusion that there are 17 such categories that are completely safe, in which no fraudulent transactions were observed.

1.1.5.2 R_emaildomain

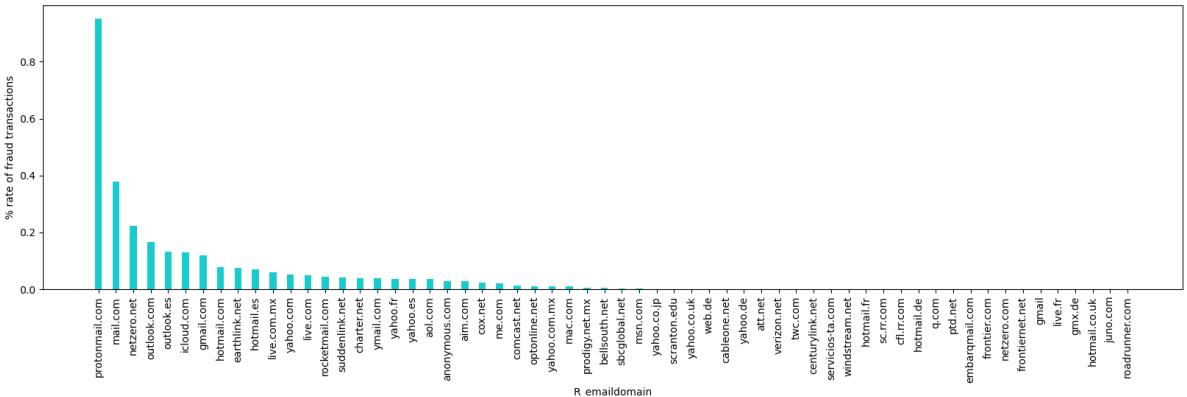
```
In [43]: plot_categorical_features(df_train, 'R_emaildomain', 'isFraud')
```

Five Number Summary

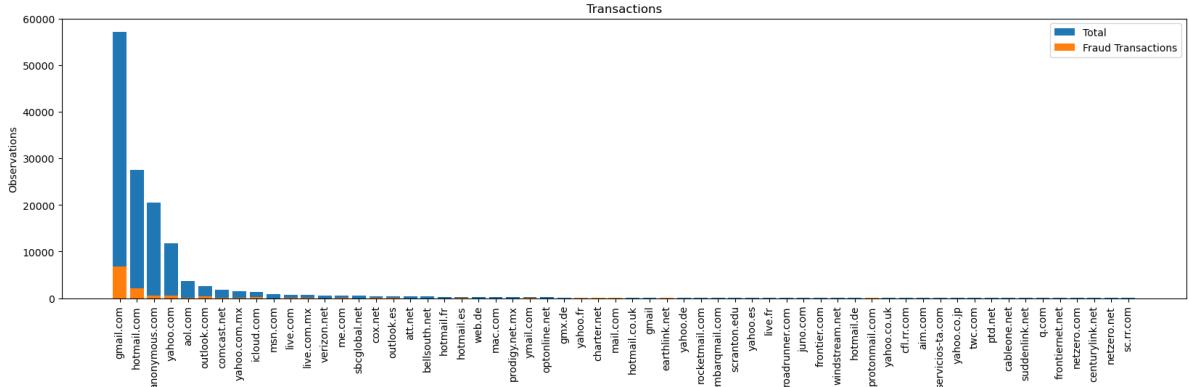
	Feature Description	Feature Description of legit transactions
	Feature Description of fraud transactions	
count	137291	126064
unique	60	60
top	gmail.com	gmail.com
freq	57147	50336
count	11227	
unique	32	
top	gmail.com	
freq	6811	
Name: R_emaildomain, dtype: object		Name: R_emaildomain, dtype: object
Name: R_emaildomain, dtype: object		

The unique values in R_emaildomain feature is: 60

R_emaildomain feature has 453249 null values, which holds 0.7675161716395164 %



1. self_case_study_1_IEEE_fraud_detection_EDA



	R_emaildomain	isFraud	total	% Occurence	Avg Fraud
16	gmail.com	6811	57147	0.096771	0.119184
19	hotmail.com	2140	27509	0.046583	0.077793
1	anonymous.com	598	20529	0.034763	0.029130
54	yahoo.com	610	11842	0.020053	0.051512
2	aol.com	129	3701	0.006267	0.034855
35	outlook.com	414	2507	0.004245	0.165138
9	comcast.net	21	1812	0.003068	0.011589
55	yahoo.com.mx	16	1508	0.002554	0.010610
23	icloud.com	180	1398	0.002367	0.128755
31	msn.com	1	852	0.001443	0.001174
25	live.com	38	762	0.001290	0.049869
26	live.com.mx	44	754	0.001277	0.058355
49	verizon.net	0	620	0.001050	0.000000
30	me.com	11	556	0.000942	0.019784
43	sbcglobal.net	1	552	0.000935	0.001812
10	cox.net	11	459	0.000777	0.023965
36	outlook.es	57	433	0.000733	0.131640
3	att.net	0	430	0.000728	0.000000
4	bellsouth.net	2	422	0.000715	0.004739
22	hotmail.fr	0	293	0.000496	0.000000
21	hotmail.es	20	292	0.000494	0.068493
50	web.de	0	237	0.000401	0.000000
28	mac.com	2	218	0.000369	0.009174
37	prodigy.net.mx	1	207	0.000351	0.004831
59	ymail.com	8	207	0.000351	0.038647
34	optonline.net	2	187	0.000317	0.010695
17	gmx.de	0	147	0.000249	0.000000
58	yahoo.fr	5	137	0.000232	0.036496
8	charter.net	5	127	0.000215	0.039370
29	mail.com	46	122	0.000207	0.377049
18	hotmail.co.uk	0	105	0.000178	0.000000
15	gmail	0	95	0.000161	0.000000
11	earthlink.net	6	79	0.000134	0.075949
56	yahoo.de	0	75	0.000127	0.000000
42	rocketmail.com	3	69	0.000117	0.043478
12	embarqmail.com	0	68	0.000115	0.000000
45	scranton.edu	0	63	0.000107	0.000000
57	yahoo.es	2	57	0.000097	0.035088
27	live.fr	0	55	0.000093	0.000000
41	roadrunner.com	0	53	0.000090	0.000000
24	juno.com	0	53	0.000090	0.000000
13	frontier.com	0	52	0.000088	0.000000
51	windstream.net	0	47	0.000080	0.000000
20	hotmail.de	0	42	0.000071	0.000000
38	protonmail.com	39	41	0.000069	0.951220
53	yahoo.co.uk	0	39	0.000066	0.000000
7	cfl.rr.com	0	37	0.000063	0.000000
0	aim.com	1	36	0.000061	0.027778
46	servicios-ta.com	0	35	0.000059	0.000000
52	yahoo.co.jp	0	33	0.000056	0.000000
48	twc.com	0	29	0.000049	0.000000
39	ptd.net	0	27	0.000046	0.000000
5	cableone.net	0	27	0.000046	0.000000
47	suddenlink.net	1	25	0.000042	0.040000
40	q.com	0	25	0.000042	0.000000
14	frontiernet.net	0	14	0.000024	0.000000
32	netzero.com	0	14	0.000024	0.000000
6	centurylink.net	0	12	0.000020	0.000000
33	netzero.net	2	9	0.000015	0.222222
44	sc.rr.com	0	8	0.000014	0.000000

Observation: This feature has maximum null values, 76.75%. The barplot revealed that a transaction made through "protonmail.com" has a greater than 95% chance of being fraudulent. "Mail.com" (37.70%) and "netzero.net" (22.22%) were the next two websites. The first two observations are comparable to the property "P_emaildomain." Furthermore, we observed that the distribution of valid transactions falls into 60 distinct categories, whereas the distribution of fraudulent transactions falls into 32 distinct categories, leading us to the conclusion that there are 28 such categories that are entirely safe, in which no fraudulent transactions were observed.

1.1.6 'M' Features

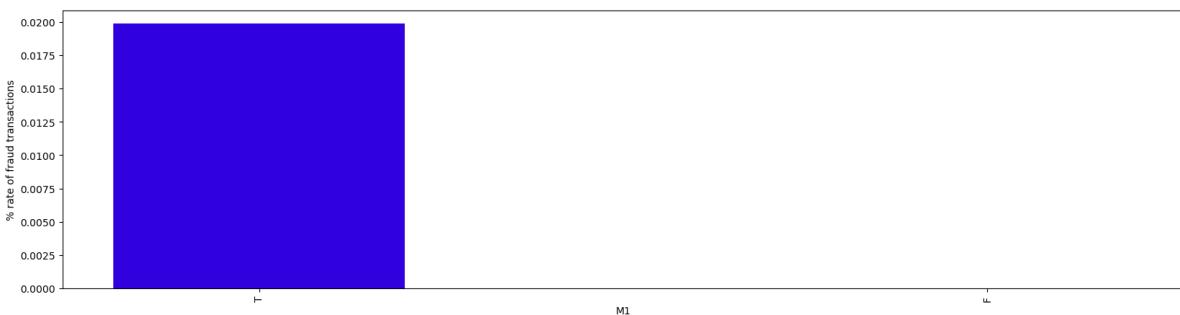
1.1.6.1 'M1'

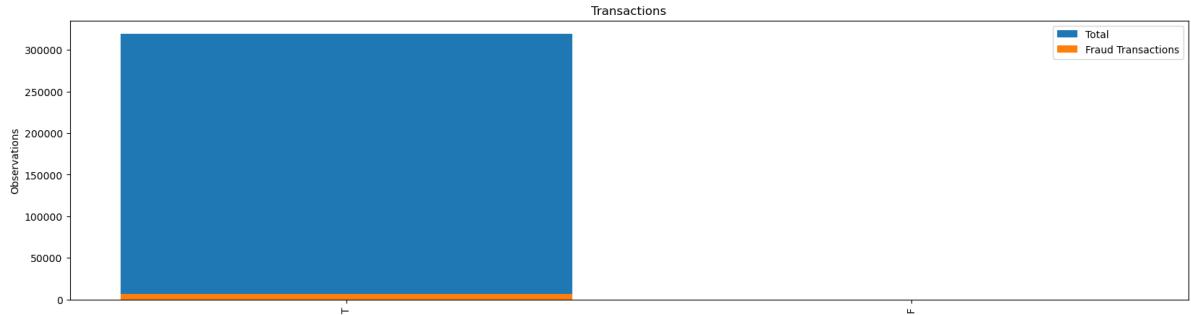
```
In [44]: plot_categorical_features(df_train, 'M1', 'isFraud')
```

Feature Description of legit transactions		Feature Description of fraud transactions	
count	319440	count	313098
unique	2	unique	2
top	T	top	T
freq	319415	freq	313073
Name: M1, dtype: object		Name: M1, dtype: object	
ame: M1, dtype: object			N

The unique values in M1 feature is: 2

M1 feature has 271100 null values, which holds 0.4590713584177194 %





M1	isFraud	total	% Occurrence	Avg Fraud
1	T	6342	319415	0.540886
0	F	0	25	0.000042

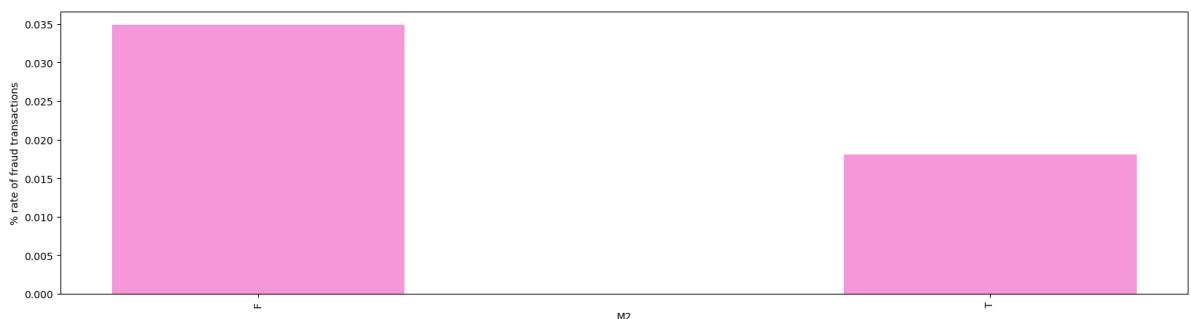
Observation: Two categories make up the entire feature. Only 25 instances of category "F" have been noticed, which is extremely infrequent (0.0%), yet out of those 25, no instances of fraudulent transactions have been noted. The major category, "T," is present 54% of the time, and the remaining values are all null.

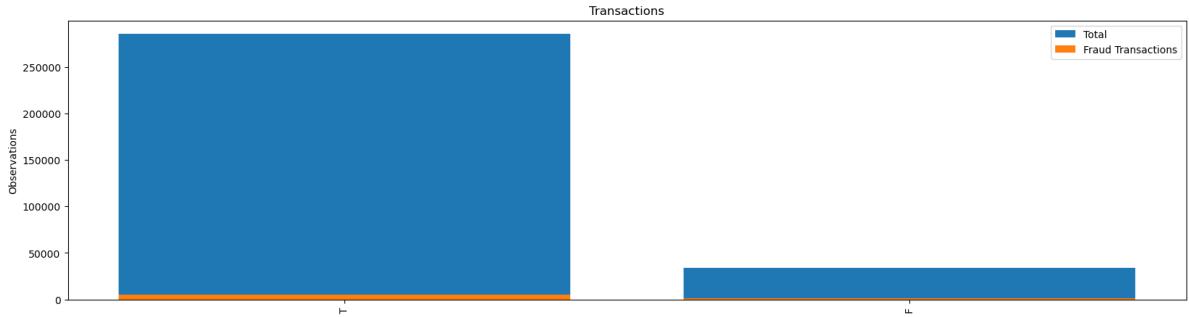
1.1.6.2 'M2'

In [45]: `plot_categorical_features(df_train, 'M2', 'isFraud')`

```
*Five Number Summary*
+-----+
| Feature Description | Feature Description of legit transactions | Feature Di
scription of fraud transactions |
+-----+
|      count    319440   |                               count    313098   |
count    6342           |                               unique     2   |
|      unique     2   |                               top       T   |
unique     2           |                               freq     280310   |
|      top       T   |                               freq    5158     |
top       T           |                               Name: M2, dtype: object |
|      freq    285468   |                               Name: M2, dtype: object |
freq    5158           |                               Name: M2, dtype: object |
|      Name: M2, dtype: object |                               Name: M2, dtype: object |
ame: M2, dtype: object |                               |
+-----+                               N
```

The unique values in M2 feature is: 2
M2 feature has 271100 null values, which holds 0.4590713584177194 %





M2	isFraud	total	% Occurrence	Avg Fraud
1	T	5158	0.483402	0.018069
0	F	1184	0.057527	0.034852

Observation: The entire feature is divided into two groups. "T" appears 48.34% of the time and dominates this trait. Only 5.75% of the time does Category "F" occur. For this feature, more cases of fraudulent transactions are offered under the "F" category than under the "T" category. The feature also contains 45.90% null values.

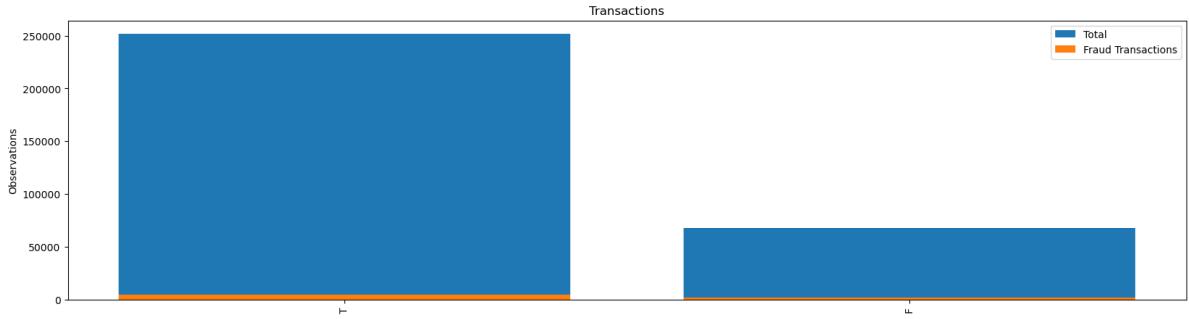
1.1.6.3 'M3'

In [46]: `plot_categorical_features(df_train, 'M3', 'isFraud')`

```
*Five Number Summary*
+-----+
| Feature Description | Feature Description of legit transactions | Feature Di
scription of fraud transactions |
+-----+
|      count      319440 |                               count      313098 |
count      6342 |                               unique      2 |
|      unique      2 |                               top        T |
unique      2 |                               freq      247438 |
|      top        T |                               freq      251731 |
top        T |                               freq      4293 |
|      freq      251731 |                               Name: M3, dtype: object |
freq      4293 |                               Name: M3, dtype: object |
| Name: M3, dtype: object |                               Name: M3, dtype: object |
ame: M3, dtype: object |                               |
+-----+                               N
```

The unique values in M3 feature is: 2
M3 feature has 271100 null values, which holds 0.4590713584177194 %





```
M3  isFraud    total % Occurence Avg Fraud
1  T      4293  251731     0.426273  0.017054
0  F      2049   67709     0.114656  0.030262
```

Observation: similar results to those of the M2 feature. and as a result, we presumptively conclude that feature M2 and M3 are correlated.

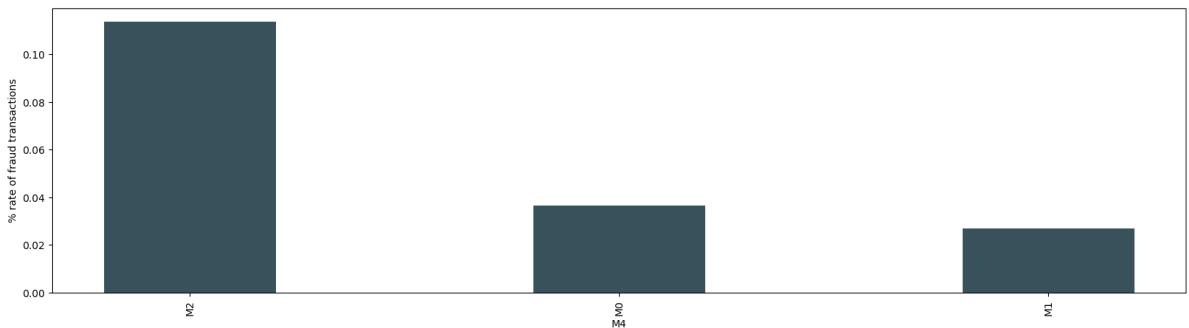
1.1.6.4 'M4'

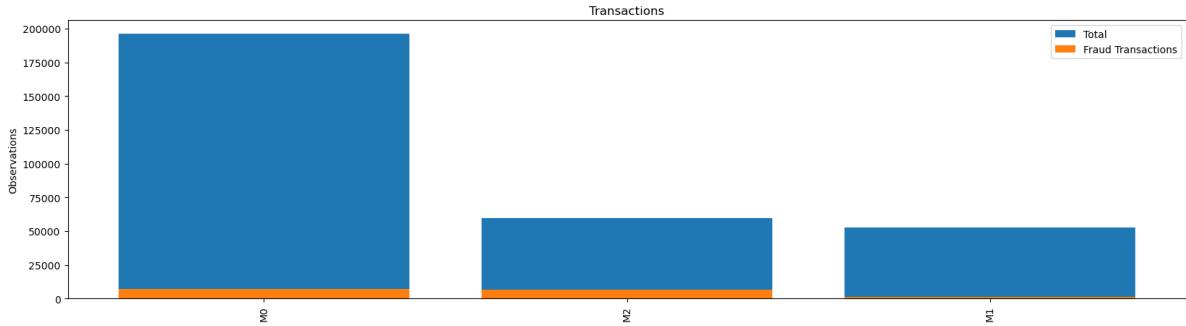
```
In [47]: plot_categorical_features(df_train, 'M4', 'isFraud')
```

```
*Five Number Summary*
+-----+
| Feature Description | Feature Description of legit transactions | Feature Di
scription of fraud transactions |
+-----+
| count      309096 |                                     count      293660
count      15436 |                                     unique      3
unique      3 |                                     top        M0
top        M0 |                                     freq      189207
freq      7198 |                                     Name: M4, dtype: object
Name: M4, dtype: object |                               Name: M4, dtype: object | N
+-----+
|                                     unique      3
top        M0 |                                     freq      189207
freq      7198 |                                     Name: M4, dtype: object
Name: M4, dtype: object |                               Name: M4, dtype: object | N
+-----+
```

The unique values in M4 feature is: 3

M4 feature has 281444 null values, which holds 0.47658753005723575 %





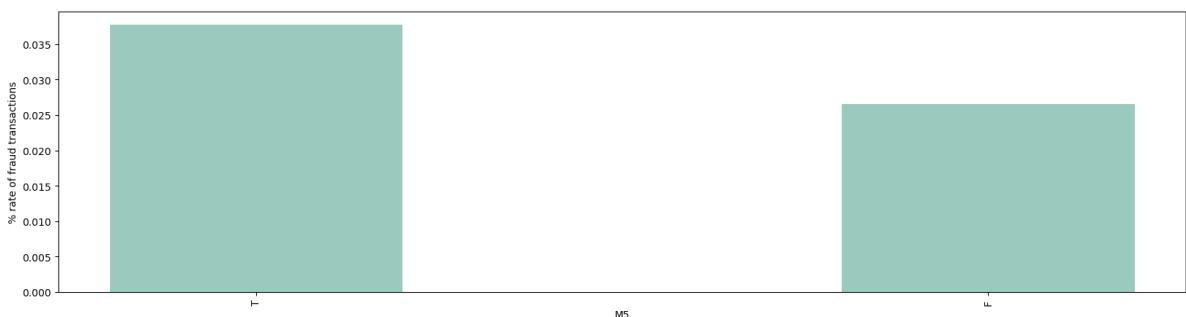
Observation: Different categories apply to feature "M4" than they do to features "M1," "M2," and "M3." It has three categories in total, with category 'M0' appearing the most frequently (33.25%), followed by categories 'M2' and 'M1'. M2 has the most fraudulent transactions out of all the categories.

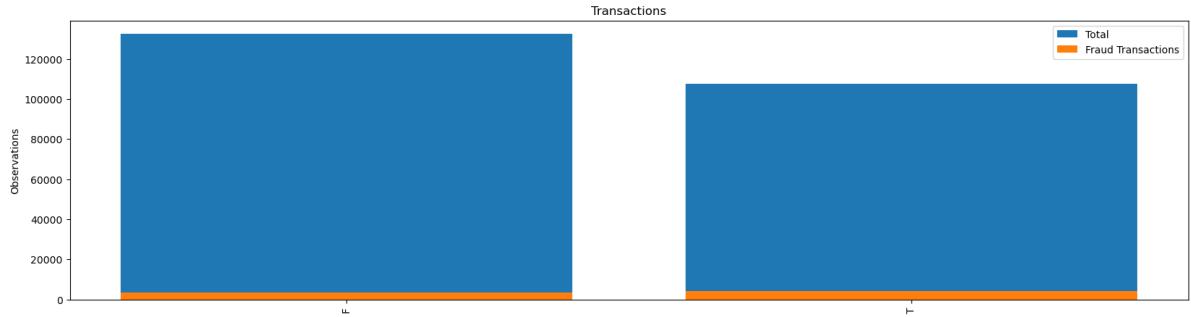
1.1.6.5 'M5'

```
In [48]: plot_categorical_features(df_train, 'M5', 'isFraud')
```

Five Number Summary		
Feature Description of legit transactions	Feature Description of legit transactions	Feature Description of legit transactions
count 240058	count 232489	
count 7569		
unique 2	unique 2	
unique 2		
top F	top F	
top T		
freq 132491	freq 128977	
freq 4055		
Name: M5, dtype: object	Name: M5, dtype: object	N
ame: M5, dtype: object		

The unique values in M5 feature is: 2
M5 feature has 350482 null values, which holds 0.5934940901547736 %





M5	isFraud	total	% Occurrence	Avg Fraud
0	F	3514	0.224356	0.026523
1	T	4055	0.182150	0.037697

Observation: Two groupings comprise the entire feature. In contrast to other features of M, the "F" category appears here the most frequently. However, when it comes to distribution of fraudulent transactions, category "T" is the most common. Here, in contrast to other aspects, the "T" category has the most fraudulent transactions. Also, 59.35% of the values in this feature are null.

1.1.6.6 'M6'

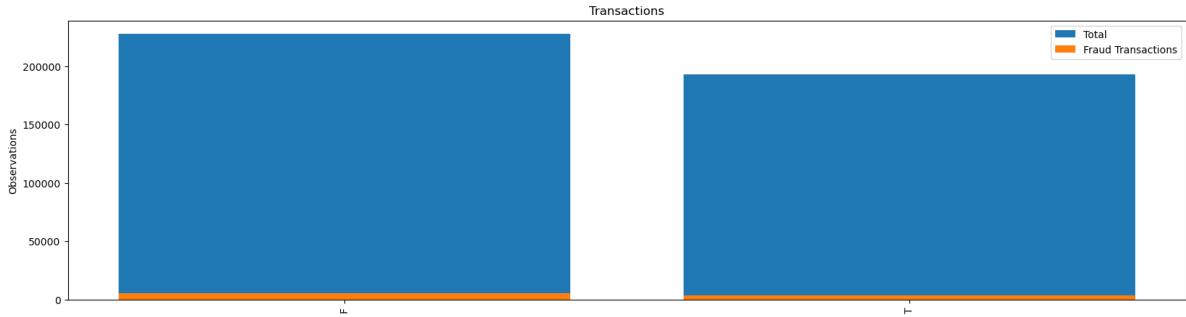
```
In [49]: plot_categorical_features(df_train, 'M6', 'isFraud')
```

Five Number Summary	
Feature Description of legit transactions	Feature Description of fraud transactions
count	count
8692	412488
unique	unique
2	2
top	top
F	F
freq	freq
227856	222459
5397	
Name: M6, dtype: object	Name: M6, dtype: object
ame: M6, dtype: object	

The unique values in M6 feature is: 2

M6 feature has 169360 null values, which holds 0.28678836319300977 %





M6	isFraud	total	% Occurrence	Avg Fraud
0	F	5397	0.385843	0.023686
1	T	3295	0.327368	0.017044

Observation: Unlike any other feature in "M," both categories "occur" in this feature in a balancing manner. The remaining findings are all comparable to feature "M5".

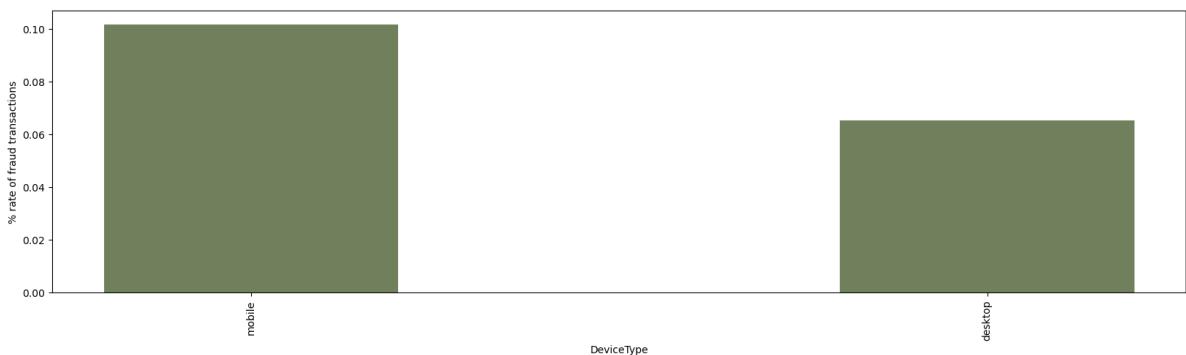
1.1.7 Device Features

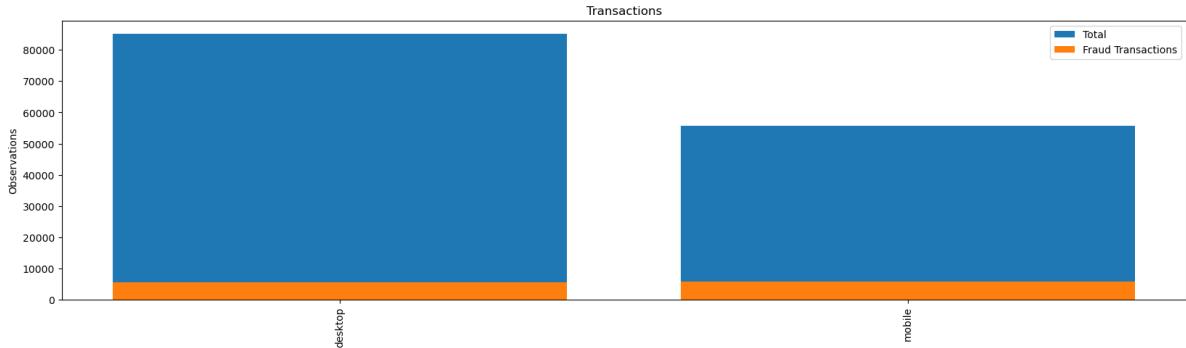
1.1.7.1 DeviceType

```
In [50]: plot_categorical_features(df_train, 'DeviceType', 'isFraud')
```

Five Number Summary		
Feature Description		Feature Description of legit transactions Feature Description of fraud transactions
count	140810	count 129599
unique	2	unique 2
top	desktop	top desktop
freq	85165	freq 79611
	Name: DeviceType, dtype: object	Name: DeviceType, dtype: object
	Name: DeviceType, dtype: object	

The unique values in DeviceType feature is: 2
 DeviceType feature has 449730 null values, which holds 0.7615572188166763 %





	DeviceType	isFraud	total	% Occurrence	Avg Fraud
0	desktop		5554	85165	0.144215
1	mobile		5657	55645	0.094227

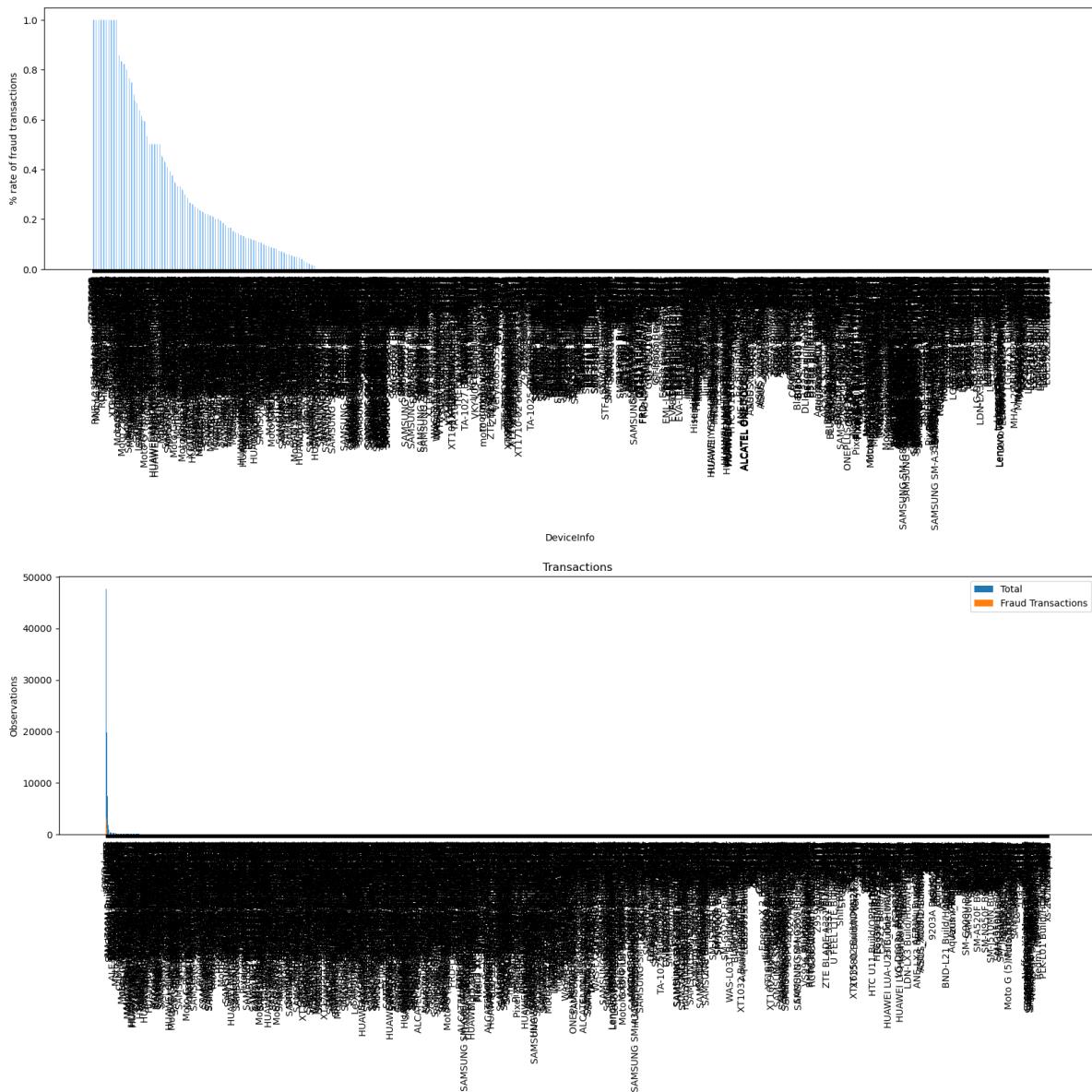
Observation: This feature has two categories: desktop and mobile. Desktop features are popping up 14.42% of the time, while mobile features occurs 9.42% of the time. The likelihood of a transaction being fraudulent increases if it is conducted from a "mobile" device. Also the feature has 76.15% of the null values.

1.1.7.2 DeviceInfo

```
In [51]: plot_categorical_features(df_train, 'DeviceInfo', 'isFraud')
```

Five Number Summary		
+-----+-----+-----+		
Feature Description Feature Description of legit transactions Fe		ature Description of fraud transactions
+-----+-----+-----+		
count	count	count
8607	118666	110059
unique	unique	unique
420	1786	1740
top	top	top
Windows	Windows	Windows
freq	freq	freq
3121	47722	44601
Name: DeviceInfo, dtype: object		Name: DeviceInfo, dtype: object
Name: DeviceInfo, dtype: object		
+-----+-----+-----+		

The unique values in DeviceInfo feature is: 1786
 DeviceInfo feature has 471874 null values, which holds 0.7990551021099332 %



	DeviceInfo	isFraud	total	% Occurence	Avg Fraud
1598	Windows	3121	47722	0.080811	0.065400
1727	iOS Device	1240	19782	0.033498	0.062683
723	MacOS	278	12573	0.021291	0.022111
1552	Trident/7.0	96	7440	0.012599	0.012903
1743	rv:11.0	76	1901	0.003219	0.039979
...
889	QwestIE8	0	1	0.000002	0.000000
892	R8106	0	1	0.000002	0.000000
1	OPJA2	0	1	0.000002	0.000000
899	RCT6K03W13	0	1	0.000002	0.000000
1785	xs-Z47b7VqTMxs	0	1	0.000002	0.000000

[1786 rows x 5 columns]

```
In [52]: # to get a dataframe with respect to corresponding feature and target variable
temp = pd.DataFrame(df_train.groupby('DeviceInfo')[['isFraud']].agg(lambda x: x.eq(1).sum()))

temp['total'] = pd.DataFrame(df_train.groupby('DeviceInfo')[['isFraud']].agg(total='sum'))
temp['% Occurence'] = temp['total'] / df_train.shape[0]
temp['Avg Fraud'] = pd.DataFrame(df_train.groupby('DeviceInfo')[['isFraud']].agg(Avg='mean'))

temp.sort_values(by=['Avg Fraud'], inplace=True, ascending=False)
temp
```

Out[52]:

	DeviceInfo	isFraud	total	% Occurrence	Avg Fraud
1716	ZTE-Z835	3	3	0.000005	1.0
235	Dream	4	4	0.000007	1.0
1448	SM-T285M	1	1	0.000002	1.0
63	A5002	1	1	0.000002	1.0
62	A466BG	3	3	0.000005	1.0
...
635	LGL62VL	0	2	0.000003	0.0
634	LGL58VL	0	2	0.000003	0.0
633	LGL57BL	0	1	0.000002	0.0
632	LGL52VL Build/LMY47V	0	2	0.000003	0.0
1785	xs-Z47b7VqTMxs	0	1	0.000002	0.0

1786 rows × 5 columns

In [53]:

```
# to find out card transactions in which fraudulent rate is more than 70% and total
cond = temp[temp['Avg Fraud'] >= 0.70]
print(cond[cond['total'] >=3])
```

1. self_case_study_1_IEEE_fraud_detection_EDA

	DeviceInfo	isFraud	total	% Occurrence	\
1716	ZTE-Z835	3	3	0.000005	
235	Dream	4	4	0.000007	
62	A466BG	3	3	0.000005	
187	Blade L3 Build/KOT49H	12	12	0.000020	
906	RNE-L23 Build/HUAWEIRNE-L23	5	5	0.000008	
1088	SAMSUNG-SM-N900A	3	3	0.000005	
803	MotoG3-TE Build/MPD24.65-33	5	5	0.000008	
810	NOKIA	10	10	0.000017	
824	Nexus 6 Build/MOB30M	11	11	0.000019	
1619	XT1033	3	3	0.000005	
315	G3123 Build/40.0.A.6.175	12	12	0.000020	
1744	rv:14.0	5	5	0.000008	
894	RCT6203W46 Build/KOT49H	9	9	0.000015	
1127	SM-A300M Build/KTU84P	3	3	0.000005	
1568	VS5012 Build/NRD90M	34	34	0.000058	
425	Hisense E51 Build/LMY47V	3	3	0.000005	
1443	SM-T217S Build/KOT49H	4	4	0.000007	
1258	SM-G925T Build/LMY47X	6	6	0.000010	
1698	Z965 Build/NMF26V	5	5	0.000008	
1689	Z835 Build/NMF26V	25	25	0.000042	
1685	Z813 Build/LMY470	11	11	0.000019	
1362	SM-J700T Build/MMB29K	8	8	0.000014	
1397	SM-N920A Build/MMB29K	21	21	0.000036	
1673	XT890 Build/9.8.2I-50_SML-25	8	8	0.000014	
1209	SM-G850M	4	4	0.000007	
65	A574BL Build/NMF26F	7	7	0.000012	
1608	XT1021	4	4	0.000007	
1725	hi6210sft Build/MRA58K	180	190	0.000322	
689	Lenovo YT3-850M Build/MMB29M	10	11	0.000019	
1542	TR10CS1 Build/JDQ39	10	11	0.000019	
804	MotoG3-TE Build/MPDS24.65-33-1-3	8	9	0.000015	
490	LG-D320 Build/KOT49I.V10a	61	71	0.000120	
1535	TA-1039 Build/NMF26F	27	32	0.000054	
452	Ilium Pad T7X Build/LMY47I	5	6	0.000010	
1714	ZTE Blade A511 Build/MRA58K	5	6	0.000010	
670	LT22i Build/6.2.A.1.100	5	6	0.000010	
1687	Z832 Build/MMB29M	5	6	0.000010	
121	Alcatel_40600 Build/MMB29M	5	6	0.000010	
789	Moto Z2 Play Build/NPSS26.118-19-11	10	12	0.000020	
1125	SM-A300H Build/LRX22G	169	203	0.000344	
719	MYA-L03 Build/HUAWEIMYA-L03	14	17	0.000029	
1534	TA-1039 Build/N2G47H	14	17	0.000029	
1565	VS425PP Build/LMY47V	23	28	0.000047	
1146	SM-A510M Build/MMB29K	45	56	0.000095	
122	Alcatel_5044R Build/NRD90M	4	5	0.000008	
667	LM-X210(G	4	5	0.000008	
1682	Z798BL Build/MMB29M	4	5	0.000008	
471	KFFOWI Build/LVY48F	51	65	0.000110	
1145	SM-A510M Build/LMY47X	14	18	0.000030	
1047	SAMSUNG SM-T710 Build/NRD90M	10	13	0.000022	
945	SAMSUNG SM-G550T1 Build/MMB29K	13	17	0.000029	
165	BLADE V8Q Build/N2G47H	9	12	0.000020	
788	Moto Z2 Play Build/NPS26.74-16-1	9	12	0.000020	
740	Moto E (4) Build/NCQ26.69-56	3	4	0.000007	
213	CRO-L03 Build/HUAWEICRO-L03	51	68	0.000115	
758	Moto G (4) Build/NPJS25.93-14-8.1-4	9	12	0.000020	
681	Lenovo PB2-670Y Build/MRA58K	3	4	0.000007	
687	Lenovo TB3-710I Build/LMY47I	11	15	0.000025	
1241	SM-G920P Build/NRD90M	39	54	0.000091	
782	Moto G Play Build/NPI26.48-36	7	10	0.000017	

Avg Fraud
1716 1.000000

```
235    1.000000
62     1.000000
187    1.000000
906    1.000000
1088   1.000000
803    1.000000
810    1.000000
824    1.000000
1619   1.000000
315    1.000000
1744   1.000000
894    1.000000
1127   1.000000
1568   1.000000
425    1.000000
1443   1.000000
1258   1.000000
1698   1.000000
1689   1.000000
1685   1.000000
1362   1.000000
1397   1.000000
1673   1.000000
1209   1.000000
65     1.000000
1608   1.000000
1725   0.947368
689    0.909091
1542   0.909091
804    0.888889
490    0.859155
1535   0.843750
452    0.833333
1714   0.833333
670    0.833333
1687   0.833333
121    0.833333
789    0.833333
1125   0.832512
719    0.823529
1534   0.823529
1565   0.821429
1146   0.803571
122    0.800000
667    0.800000
1682   0.800000
471    0.784615
1145   0.777778
1047   0.769231
945    0.764706
165    0.750000
788    0.750000
740    0.750000
213    0.750000
758    0.750000
681    0.750000
687    0.733333
1241   0.722222
782    0.700000
```

```
In [54]: nfr = len(cond[cond['total'] >=3])
print('Number of cases where transaction been made more than thrice and chances of
```

Number of cases where transaction been made more than thrice and chances of that being fraud are more than 70%: 60

```
In [55]: print('Number of card transactions when the likelihood of fraud is 100 percent:',  
      print('Number of card transactions where there are more than 70% fraud cases:', len(  
      print(f'There are {nfr} such cases where card transaction occurred three or more than three times and the chances of that being fraud are more than 70%'))
```

Number of card transactions when the likelihood of fraud is 100 percent: 46

Number of card transactions where there are more than 70% fraud cases: 79

There are 60 such cases where card transaction occurred three or more than three times and the chances of that being fraud are more than 70%

Observation: This feature has 1786 unique categories with 79.90% of null values. 'windows' is the most frequently occurring category at the rate 8%.

1.1.8 ID Features

```
In [56]: for fea in cat_fea:  
    if fea.startswith('id'):  
        print(f'Analysis of {fea} feature')  
        plot_categorical_features(df_train, fea, 'isFraud')  
        print('\n')  
        print('*'*120)  
        print('\n')
```

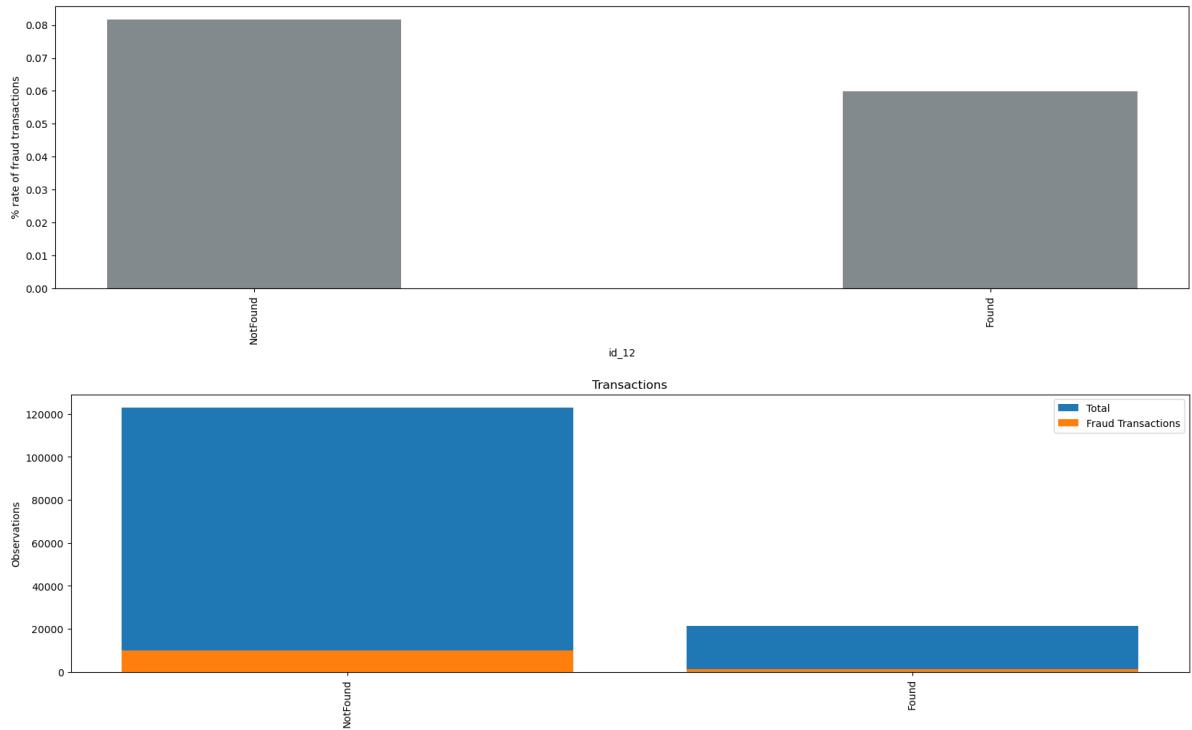
Analysis of id_12 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	144233	count	132915		
unique	11318	unique	2		
top	NotFound	top	NotFound		
freq	123025	freq	112976		
Name: id_12, dtype: object		Name: id_12, dtype: object			
Name: id_12, dtype: object					

The unique values in id_12 feature is: 2

id_12 feature has 446307 null values, which holds 0.7557608290716971 %



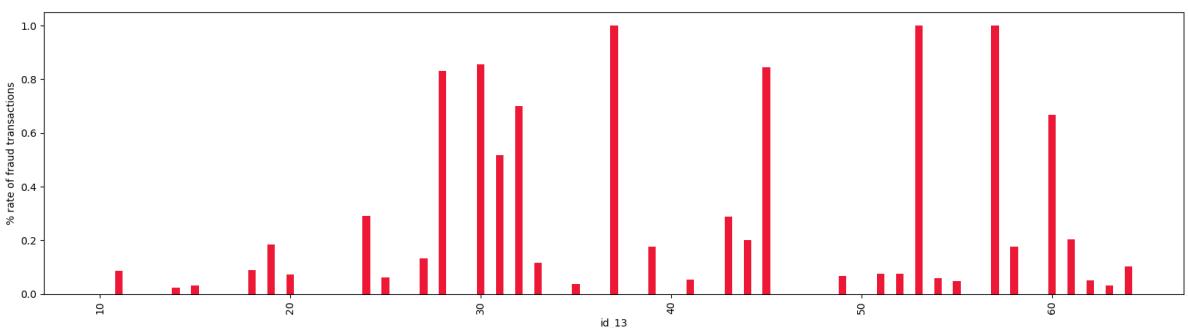
	id_12	isFraud	total	% Occurence	Avg Fraud
1	NotFound	10049	123025	0.208326	0.081683
0	Found	1269	21208	0.035913	0.059836

Analysis of id_13 feature

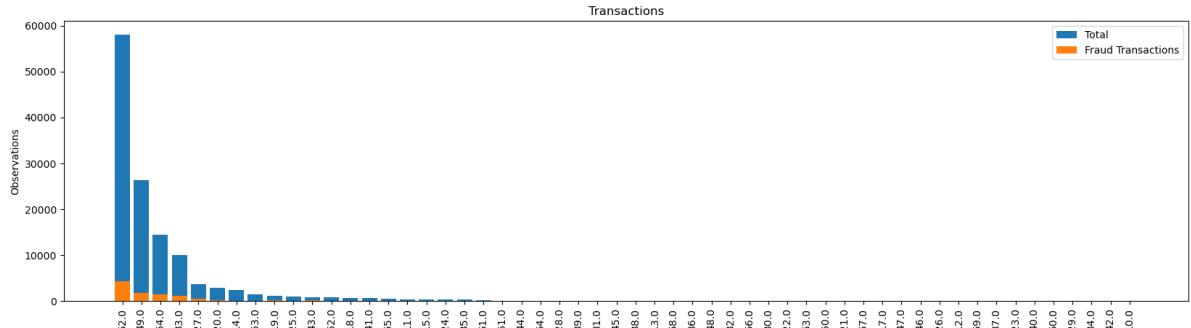
Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	127320.000000	count	116732.000000
mean	48.053071	mean	48.122357
std	11.774858	std	11.722255
min	10.000000	min	10.000000
25%	49.000000	25%	49.000000
50%	52.000000	50%	52.000000
75%	52.000000	75%	52.000000
max	64.000000	max	64.000000
	Name: id_13, dtype: float64	Name: id_13, dtype: float64	Name: id_13, dtype: float64

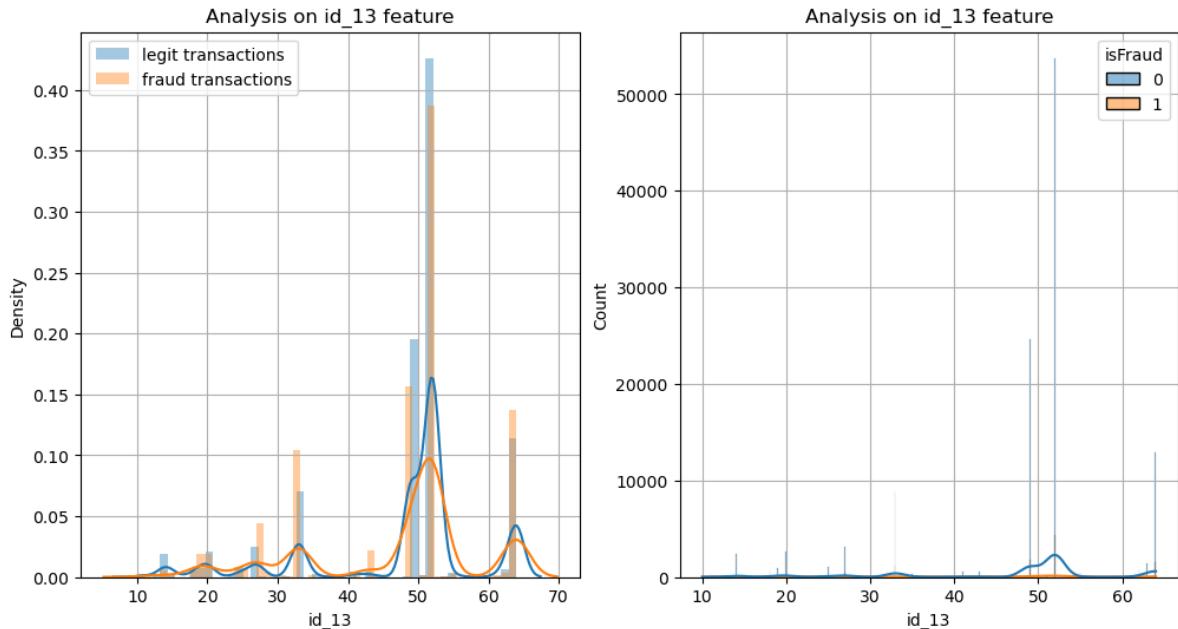
The unique values in id_13 feature is: 54
 id_13 feature has 463220 null values, which holds 0.7844007179869272 %



1. self_case_study_1_IEEE_fraud_detection_EDA



	id_13	isFraud	total	% Occurence	Avg Fraud
41	52.0	4343	58099	0.098383	0.074752
38	49.0	1761	26365	0.044646	0.066793
53	64.0	1491	14429	0.024434	0.103334
22	33.0	1172	10048	0.017015	0.116640
16	27.0	492	3666	0.006208	0.134206
9	20.0	212	2878	0.004874	0.073662
4	14.0	59	2499	0.004232	0.023609
52	63.0	49	1468	0.002486	0.033379
8	19.0	212	1147	0.001942	0.184830
14	25.0	67	1066	0.001805	0.062852
32	43.0	243	842	0.001426	0.288599
51	62.0	41	813	0.001377	0.050431
7	18.0	61	688	0.001165	0.088663
30	41.0	35	654	0.001107	0.053517
44	55.0	25	523	0.000886	0.047801
1	11.0	32	373	0.000632	0.085791
5	15.0	12	372	0.000630	0.032258
13	24.0	93	318	0.000538	0.292453
24	35.0	12	317	0.000537	0.037855
40	51.0	15	196	0.000332	0.076531
50	61.0	22	108	0.000183	0.203704
33	44.0	17	85	0.000144	0.200000
43	54.0	4	67	0.000113	0.059701
17	28.0	49	59	0.000100	0.830508
28	39.0	6	34	0.000058	0.176471
20	31.0	15	29	0.000049	0.517241
34	45.0	22	26	0.000044	0.846154
27	38.0	0	22	0.000037	0.000000
3	13.0	0	21	0.000036	0.000000
47	58.0	3	17	0.000029	0.176471
25	36.0	0	15	0.000025	0.000000
37	48.0	0	10	0.000017	0.000000
21	32.0	7	10	0.000017	0.700000
45	56.0	0	9	0.000015	0.000000
19	30.0	6	7	0.000012	0.857143
11	22.0	0	5	0.000008	0.000000
42	53.0	4	4	0.000007	1.000000
49	60.0	2	3	0.000005	0.666667
10	21.0	0	3	0.000005	0.000000
46	57.0	3	3	0.000005	1.000000
6	17.0	0	3	0.000005	0.000000
36	47.0	0	3	0.000005	0.000000
35	46.0	0	3	0.000005	0.000000
15	26.0	0	2	0.000003	0.000000
2	12.0	0	2	0.000003	0.000000
48	59.0	0	1	0.000002	0.000000
26	37.0	1	1	0.000002	1.000000
12	23.0	0	1	0.000002	0.000000
29	40.0	0	1	0.000002	0.000000
39	50.0	0	1	0.000002	0.000000
18	29.0	0	1	0.000002	0.000000
23	34.0	0	1	0.000002	0.000000
31	42.0	0	1	0.000002	0.000000
0	10.0	0	1	0.000002	0.000000



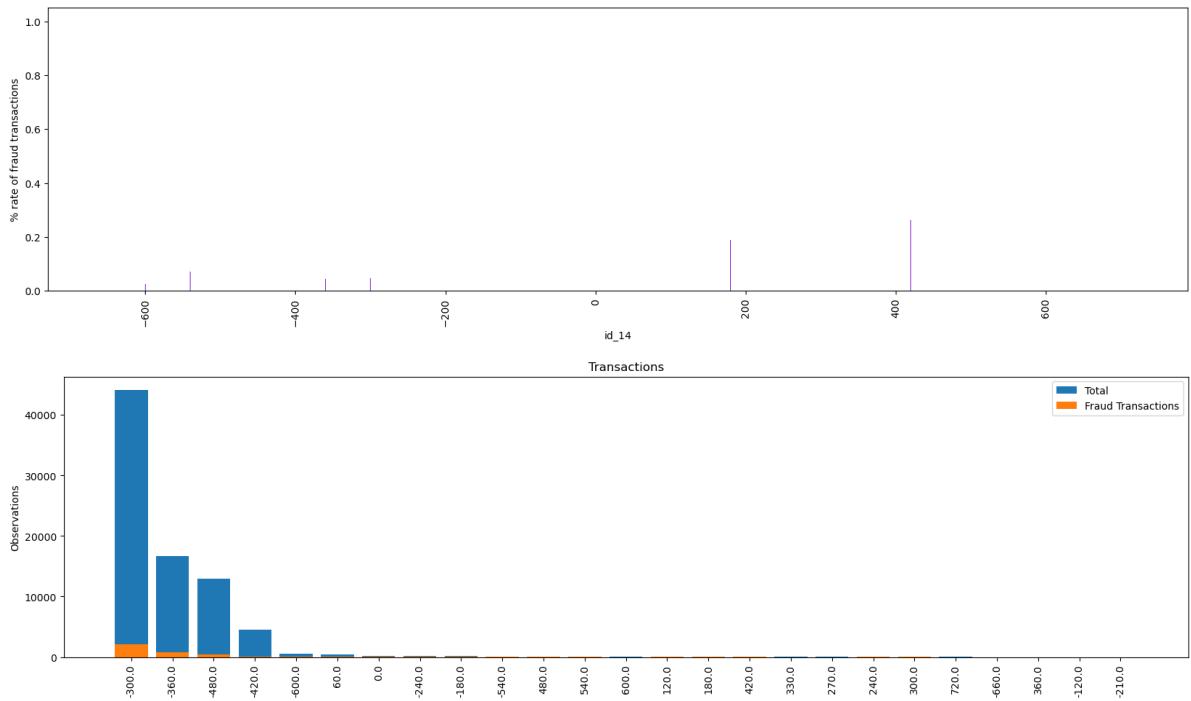
```
*****
*****
```

Analysis of id_14 feature

Five Number Summary

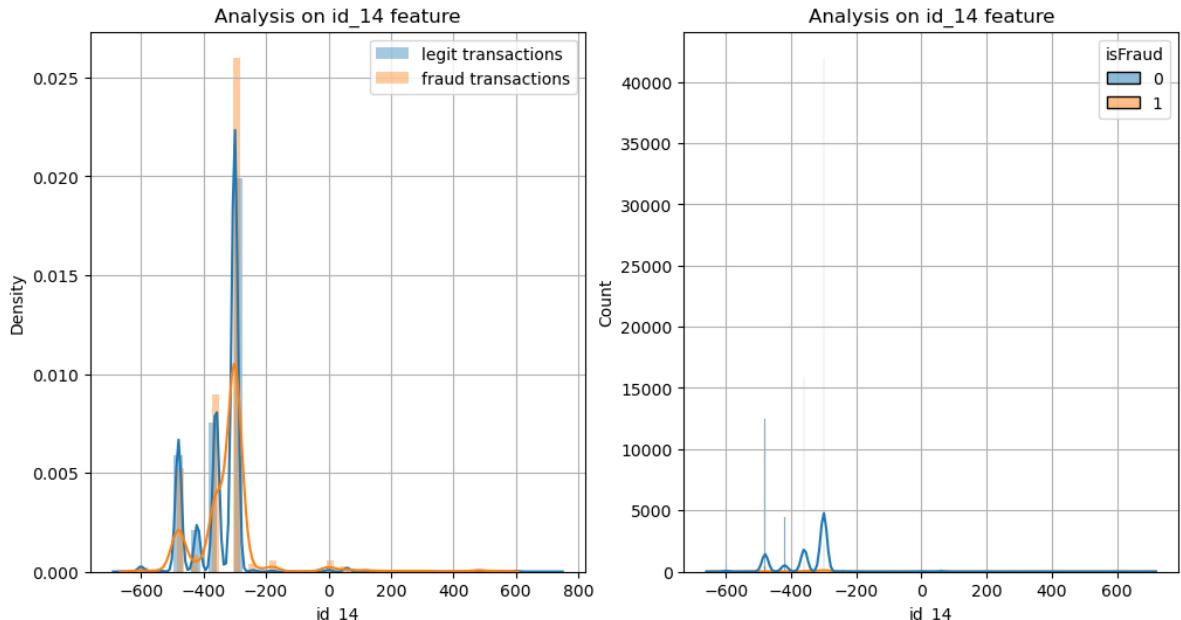
	Feature Description of legit transactions	Feature Description of fraud transactions
count	76465.000000	3579.000000
mean	-345.669130	-344.507146
std	92.321506	93.695502
min	-660.000000	-600.000000
25%	-360.000000	-360.000000
50%	-300.000000	-300.000000
75%	-300.000000	-300.000000
max	720.000000	540.000000
Name:	id_14, dtype: float64	id_14, dtype: float64

The unique values in id_14 feature is: 25
 id_14 feature has 510496 null values, which holds 0.8644562603718631 %



	id_14	isFraud	total	% Occurrence	Avg Fraud
6	-300.0	2124	44121	0.074713	0.048140
5	-360.0	734	16661	0.028213	0.044055
3	-480.0	426	12891	0.021829	0.033046
4	-420.0	72	4542	0.007691	0.015852
1	-600.0	12	498	0.000843	0.024096
12	60.0	21	369	0.000625	0.056911
11	0.0	45	192	0.000325	0.234375
7	-240.0	31	159	0.000269	0.194969
9	-180.0	48	126	0.000213	0.380952
2	-540.0	8	111	0.000188	0.072072
21	480.0	12	80	0.000135	0.150000
22	540.0	7	64	0.000108	0.109375
23	600.0	0	62	0.000105	0.000000
13	120.0	13	41	0.000069	0.317073
14	180.0	7	37	0.000063	0.189189
20	420.0	5	19	0.000032	0.263158
18	330.0	0	17	0.000029	0.000000
16	270.0	0	15	0.000025	0.000000
15	240.0	6	13	0.000022	0.461538
17	300.0	7	12	0.000020	0.583333
24	720.0	0	9	0.000015	0.000000
0	-660.0	0	2	0.000003	0.000000
19	360.0	1	1	0.000002	1.000000
10	-120.0	0	1	0.000002	0.000000
8	-210.0	0	1	0.000002	0.000000

1. self_case_study_1_IEEE_fraud_detection_EDA



```
*****
*****
```

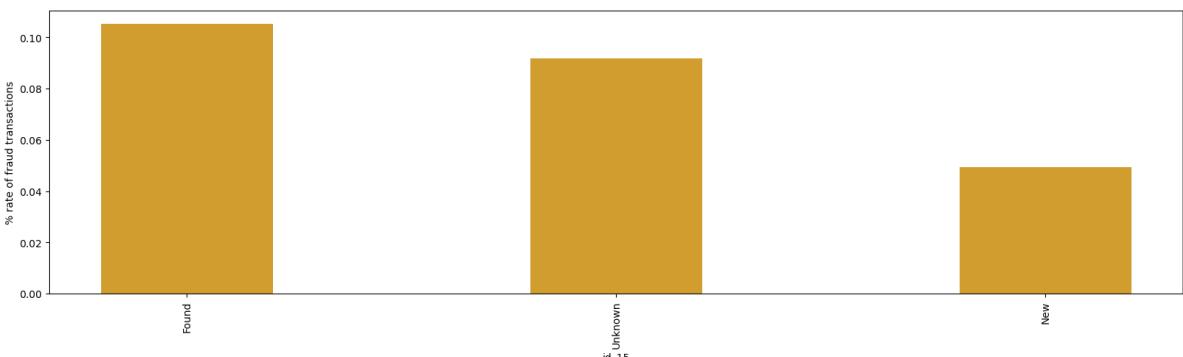
Analysis of id_15 feature

Five Number Summary

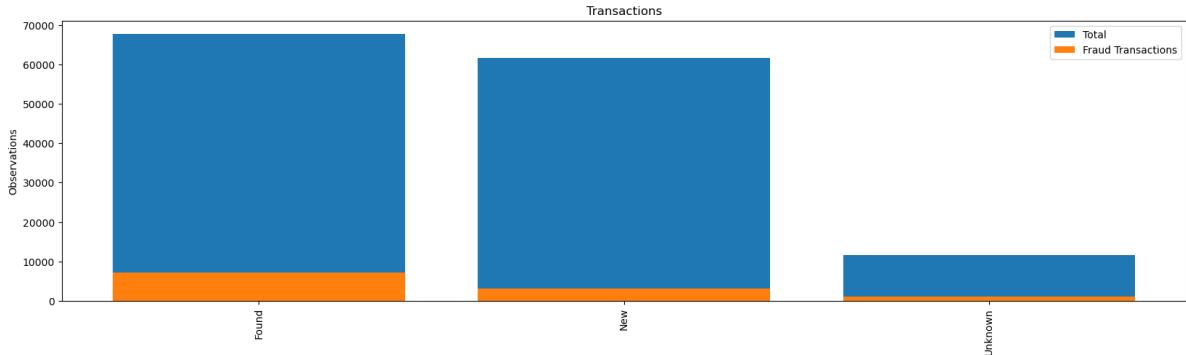
Feature Description of fraud transactions	Feature Description of legit transactions	Feature
count 140985	count 129763	
count 11222	unique 3	
unique 3	top Found	
top Found	freq 67728	
freq 7118	freq 60610	
Name: id_15, dtype: object	Name: id_15, dtype: object	
Name: id_15, dtype: object		

The unique values in id_15 feature is: 3

id_15 feature has 449555 null values, which holds 0.7612608798726589 %



1. self_case_study_1_IEEE_fraud_detection_EDA



	id_15	isFraud	total	% Occurrence	Avg Fraud
0	Found		7118	67728	0.114688 0.105097
1	New		3034	61612	0.104332 0.049244
2	Unknown		1070	11645	0.019719 0.091885

Analysis of id_16 feature

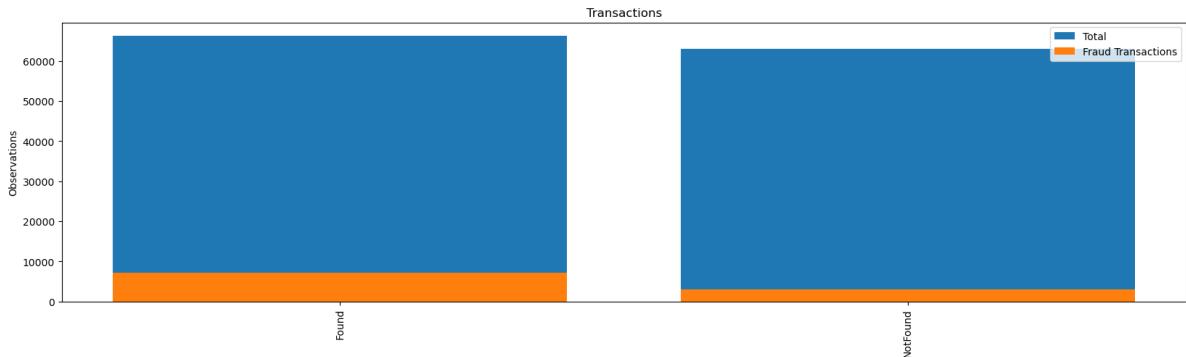
Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	129340	count	119188	
unique	10152	unique	2	
freq	66324	freq	59976	
top	Found	top	NotFound	
Name: id_16, dtype: object		Name: id_16, dtype: object		
Name: id_16, dtype: object				

The unique values in id_16 feature is: 2

id_16 feature has 461200 null values, which holds 0.7809801198902699 %





	id_16	isFraud	total	% Occurence	Avg Fraud
0	Found		7112	66324	0.112311 0.107231
1	NotFound		3040	63016	0.106709 0.048242

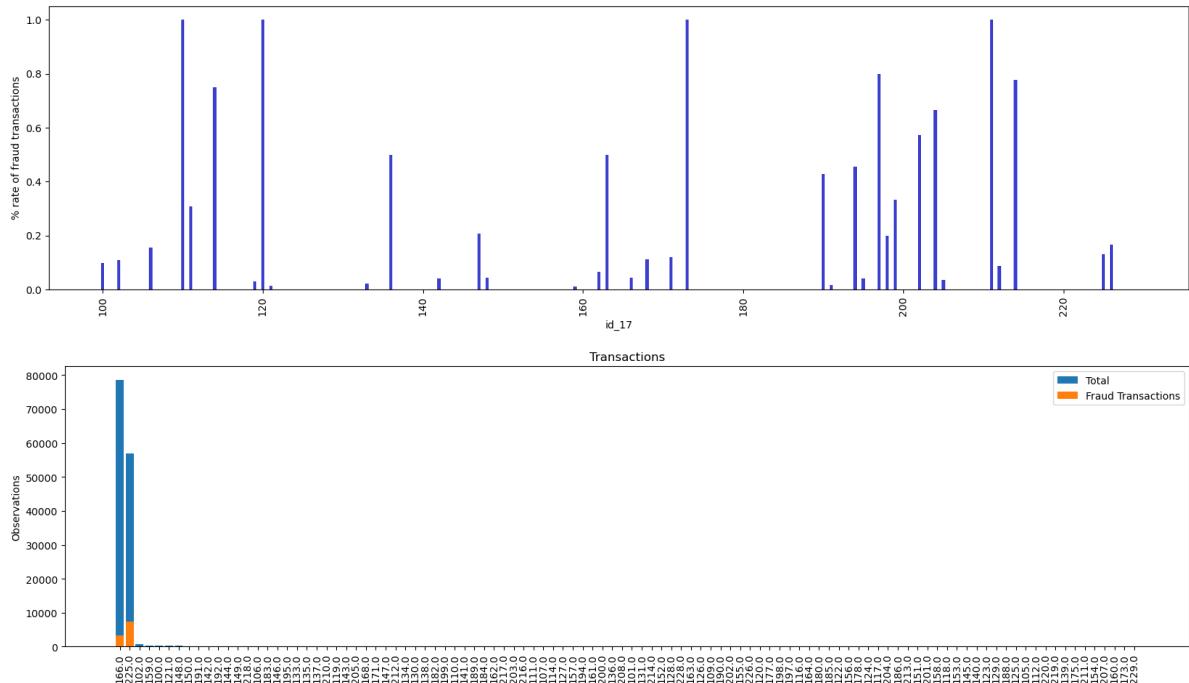
Analysis of id_17 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139369.000000	count	128323.000000	count
mean	189.451377	mean	188.113697	mean
std	30.375360	std	30.049242	std
min	100.000000	min	100.000000	min
25%	166.000000	25%	166.000000	25%
50%	166.000000	50%	166.000000	50%
75%	225.000000	75%	225.000000	75%
max	229.000000	max	229.000000	max
	Name: id_17, dtype: float64		Name: id_17, dtype: float64	
	Name: id_17, dtype: float64			

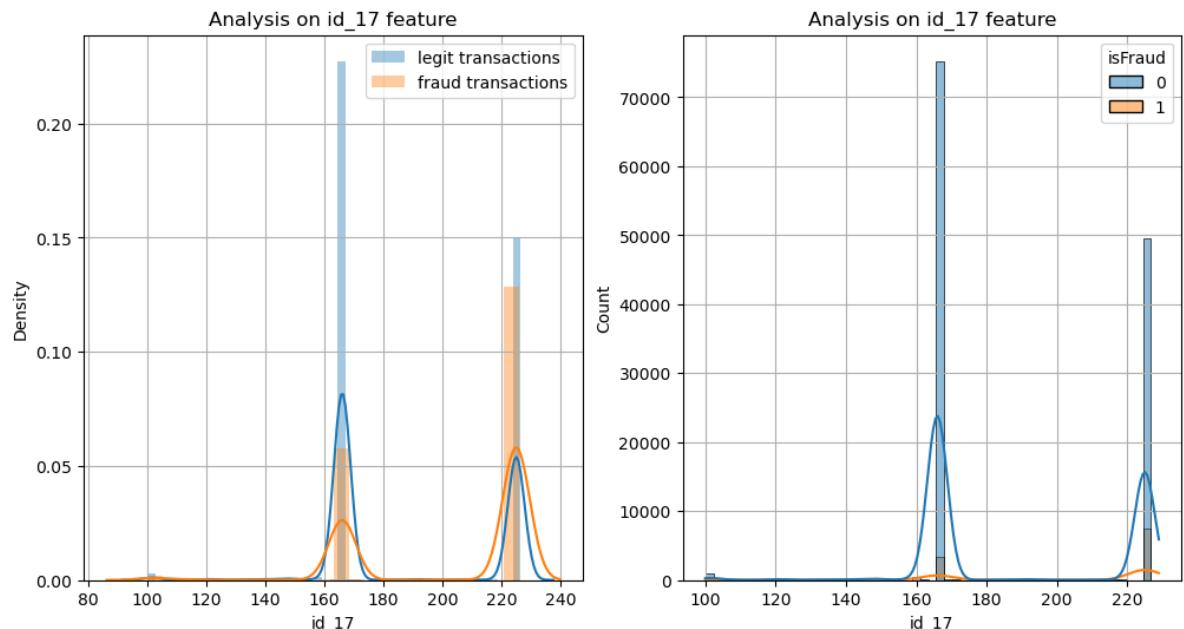
The unique values in id_17 feature is: 104

id_17 feature has 451171 null values, which holds 0.7639973583499847 %



	id_17	isFraud	total	% Occurrence	Avg Fraud
59	166.0	3350	78631	0.133151	0.042604
100	225.0	7452	56968	0.096468	0.130810
2	102.0	76	689	0.001167	0.110305
53	159.0	4	352	0.000596	0.011364
0	100.0	33	336	0.000569	0.098214
..
48	154.0	0	1	0.000002	0.000000
88	207.0	0	1	0.000002	0.000000
54	160.0	0	1	0.000002	0.000000
62	173.0	1	1	0.000002	1.000000
103	229.0	0	1	0.000002	0.000000

[104 rows x 5 columns]



```
*****
*****
```

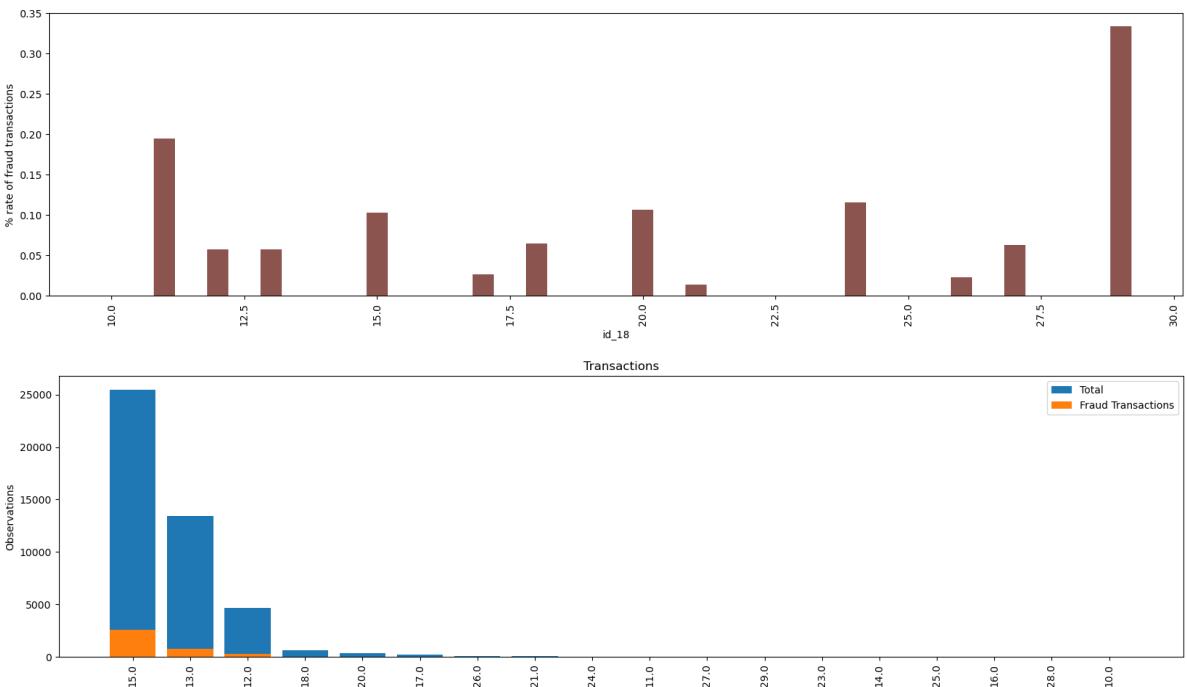
Analysis of id_18 feature

Five Number Summary

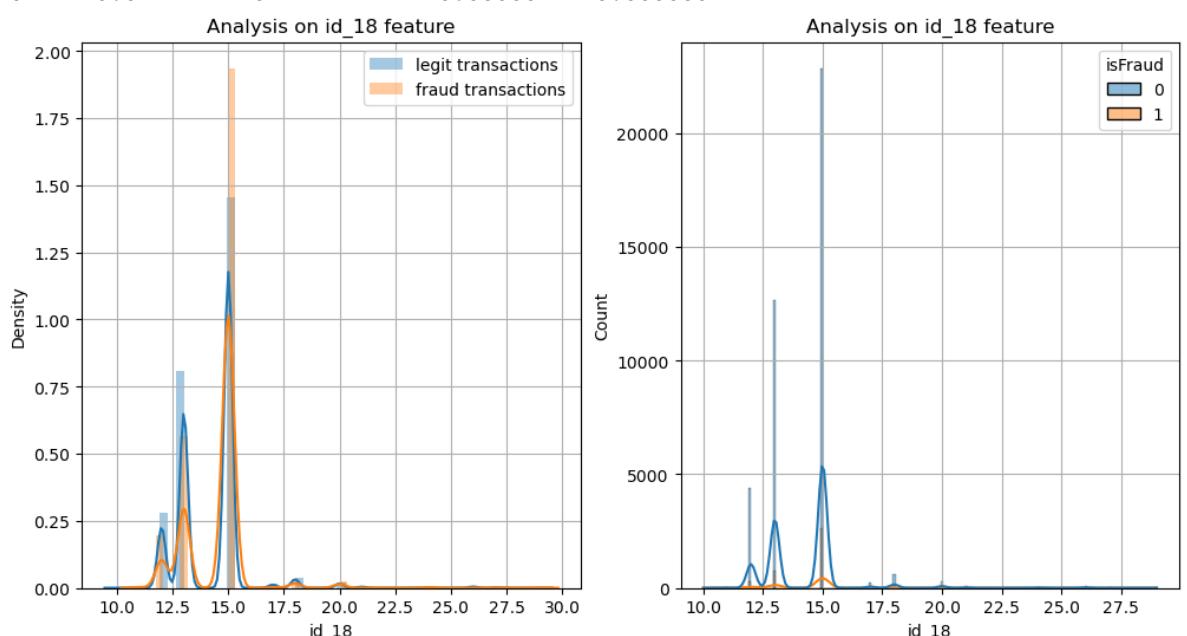
Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	45113.000000	count	41378.000000	count
mean	14.237337	mean	14.213882	mean
std	1.561302	std	1.571254	std
min	10.000000	min	10.000000	min
25%	11.000000	25%	13.000000	25%
50%	13.000000	50%	15.000000	50%
75%	15.000000	75%	15.000000	75%
max	29.000000	max	29.000000	max
Name: id_18, dtype: float64		Name: id_18, dtype: float64		Name: id_18, dtype: float64

The unique values in id_18 feature is: 18

id_18 feature has 545427 null values, which holds 0.9236072069631185 %



	id_18	isFraud	total	% Occurence	Avg Fraud
5	15.0	2605	25489	0.043162	0.102201
3	13.0	759	13439	0.022757	0.056477
2	12.0	266	4656	0.007884	0.057131
8	18.0	42	650	0.001101	0.064615
9	20.0	36	339	0.000574	0.106195
7	17.0	6	233	0.000395	0.025751
14	26.0	2	89	0.000151	0.022472
10	21.0	1	78	0.000132	0.012821
12	24.0	6	52	0.000088	0.115385
1	11.0	7	36	0.000061	0.194444
15	27.0	2	32	0.000054	0.062500
17	29.0	3	9	0.000015	0.333333
11	23.0	0	4	0.000007	0.000000
4	14.0	0	3	0.000005	0.000000
13	25.0	0	1	0.000002	0.000000
6	16.0	0	1	0.000002	0.000000
16	28.0	0	1	0.000002	0.000000
0	10.0	0	1	0.000002	0.000000



```
*****
*****
```

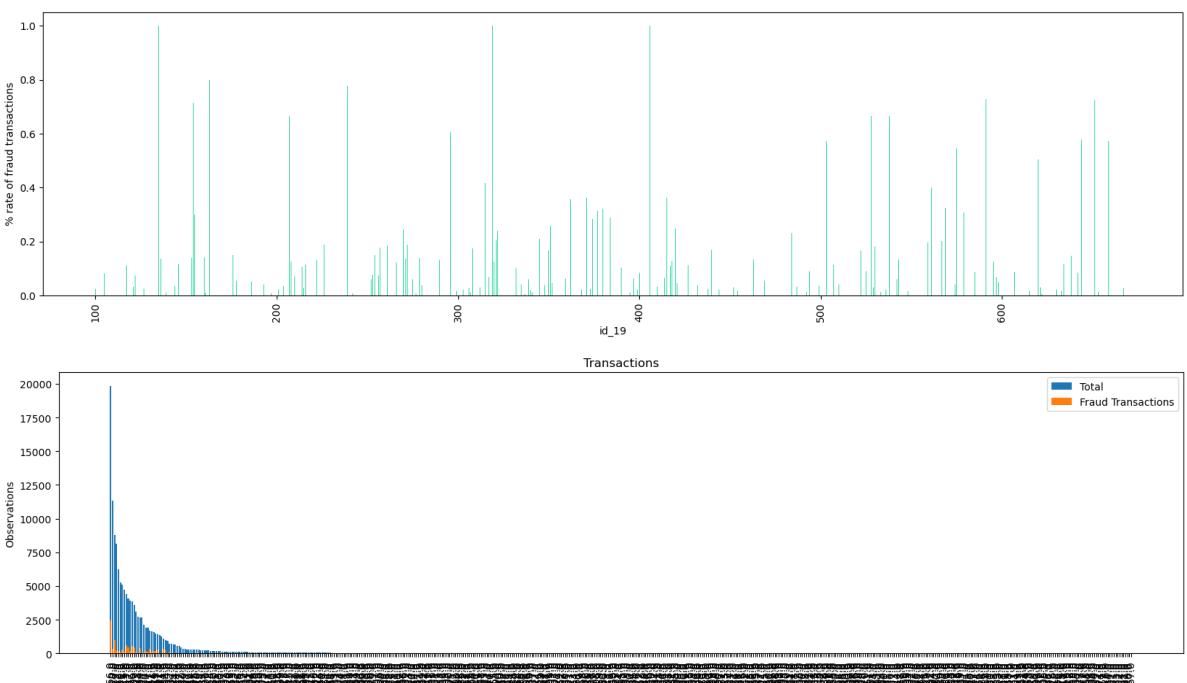
Analysis of id_19 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139318.000000	count	128280.000000	
count	11038.000000			
mean	353.128174	mean	354.854950	
mean	333.060156			
std	141.095343	std	142.178953	
std	126.113487			
min	100.000000	min	100.000000	
min	100.000000			
25%	266.000000	25%	266.000000	
25%	266.000000			
50%	341.000000	50%	341.000000	
50%	290.000000			
75%	427.000000	75%	427.000000	
75%	417.000000			
max	671.000000	max	671.000000	
max	667.000000			
Name: id_19, dtype: float64		Name: id_19, dtype: float64		
Name: id_19, dtype: float64				

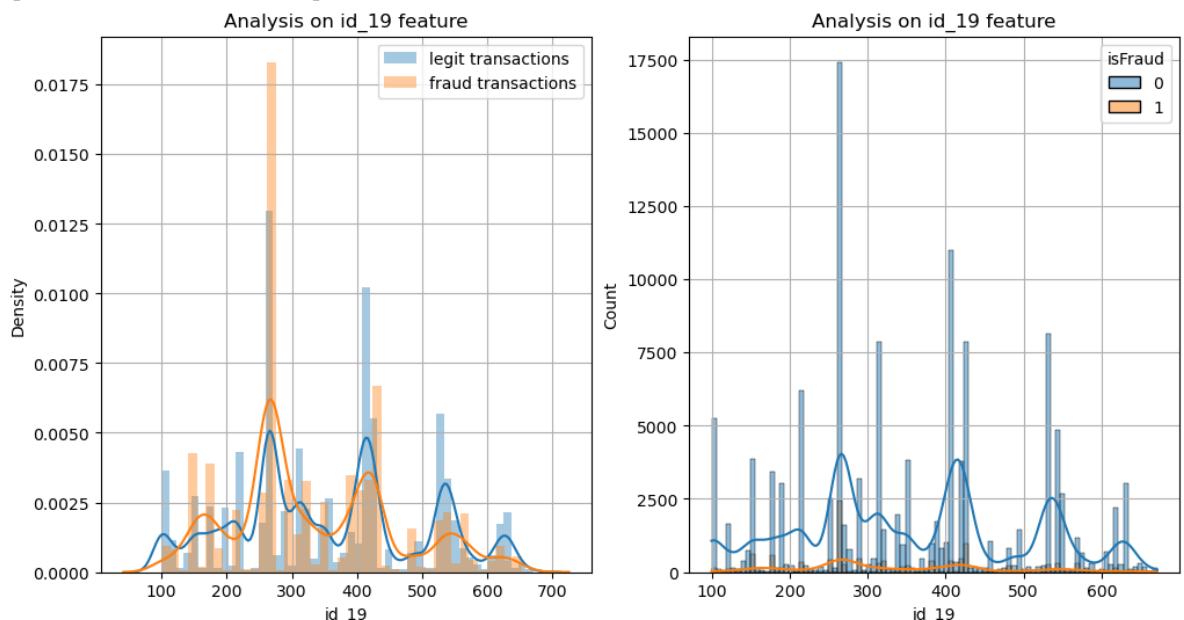
The unique values in id_19 feature is: 522

id_19 feature has 451222 null values, which holds 0.7640837199850984 %



	id_19	isFraud	total	% Occurrence	Avg Fraud
150	266.0	2447	19849	0.033612	0.123281
282	410.0	367	11318	0.019166	0.032426
298	427.0	993	8808	0.014915	0.112738
389	529.0	251	8122	0.013754	0.030904
193	312.0	178	6227	0.010545	0.028585
..
83	191.0	0	1	0.000002	0.000000
241	367.0	0	1	0.000002	0.000000
72	179.0	0	1	0.000002	0.000000
260	386.0	0	1	0.000002	0.000000
334	467.0	0	1	0.000002	0.000000

[522 rows x 5 columns]



```
*****
*****
```

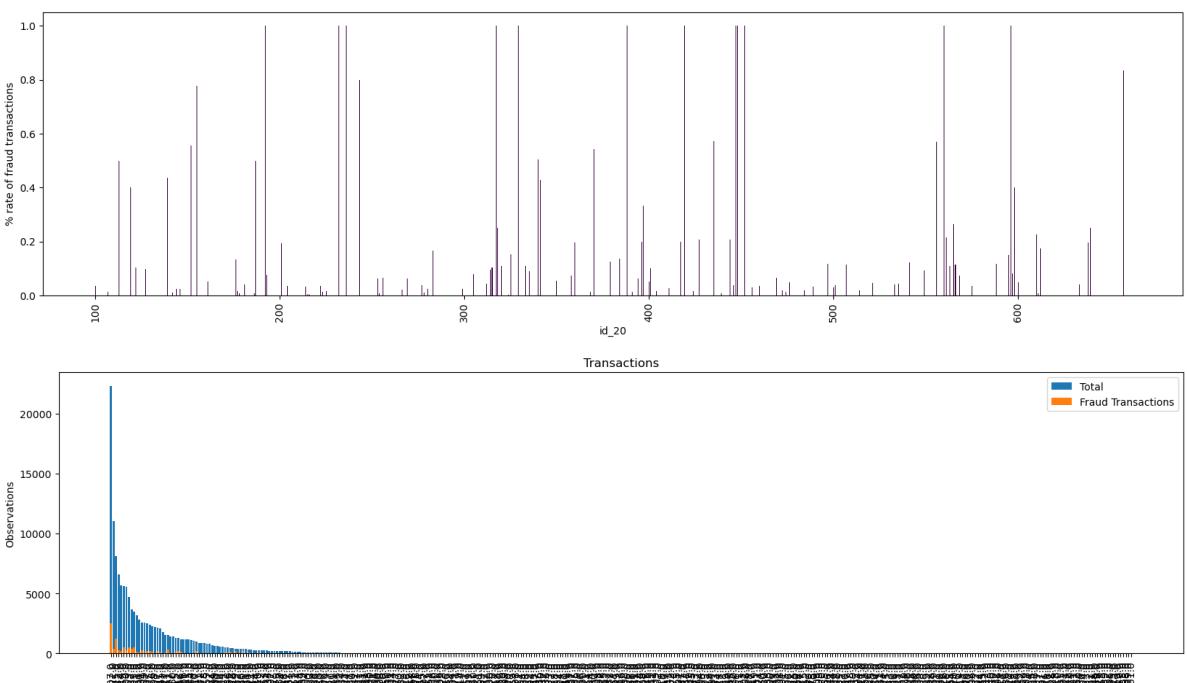
Analysis of id_20 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139261.000000	count	128230.000000	
count	11031.000000			
mean	403.882666	mean	401.133705	
mean	435.838002			
std	152.160327	std	152.826127	
std	140.303630			
min	100.000000	min	100.000000	
min	100.000000			
25%	256.000000	25%	256.000000	
25%	325.000000			
50%	472.000000	50%	469.000000	
50%	507.000000			
75%	533.000000	75%	533.000000	
75%	549.000000			
max	661.000000	max	661.000000	
max	657.000000			
Name: id_20, dtype: float64		Name: id_20, dtype: float64		
Name: id_20, dtype: float64				

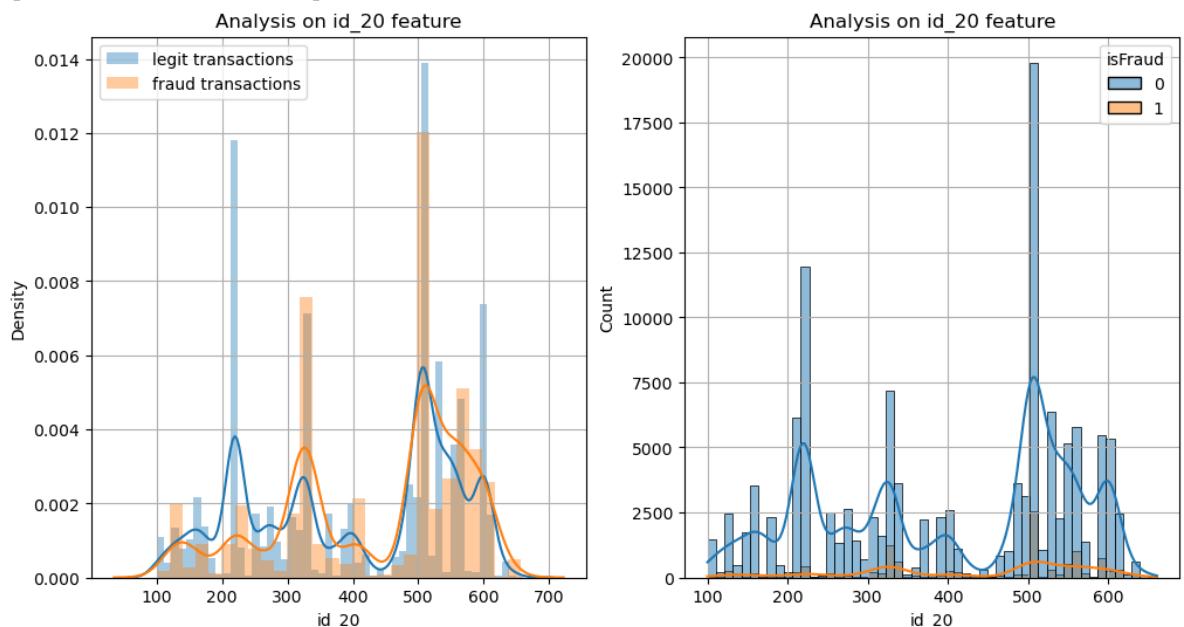
The unique values in id_20 feature is: 394

id_20 feature has 451279 null values, which holds 0.7641802418125783 %



	id_20	isFraud	total	% Occurrence	Avg Fraud
293	507.0	2542	22311	0.037781	0.113935
83	222.0	390	11065	0.018737	0.035246
160	325.0	1246	8133	0.013772	0.153203
307	533.0	278	6611	0.011195	0.042051
77	214.0	188	5664	0.009591	0.033192
..
124	279.0	0	1	0.000002	0.000000
128	285.0	0	1	0.000002	0.000000
130	288.0	0	1	0.000002	0.000000
131	289.0	0	1	0.000002	0.000000
393	661.0	0	1	0.000002	0.000000

[394 rows x 5 columns]



```
*****
*****
```

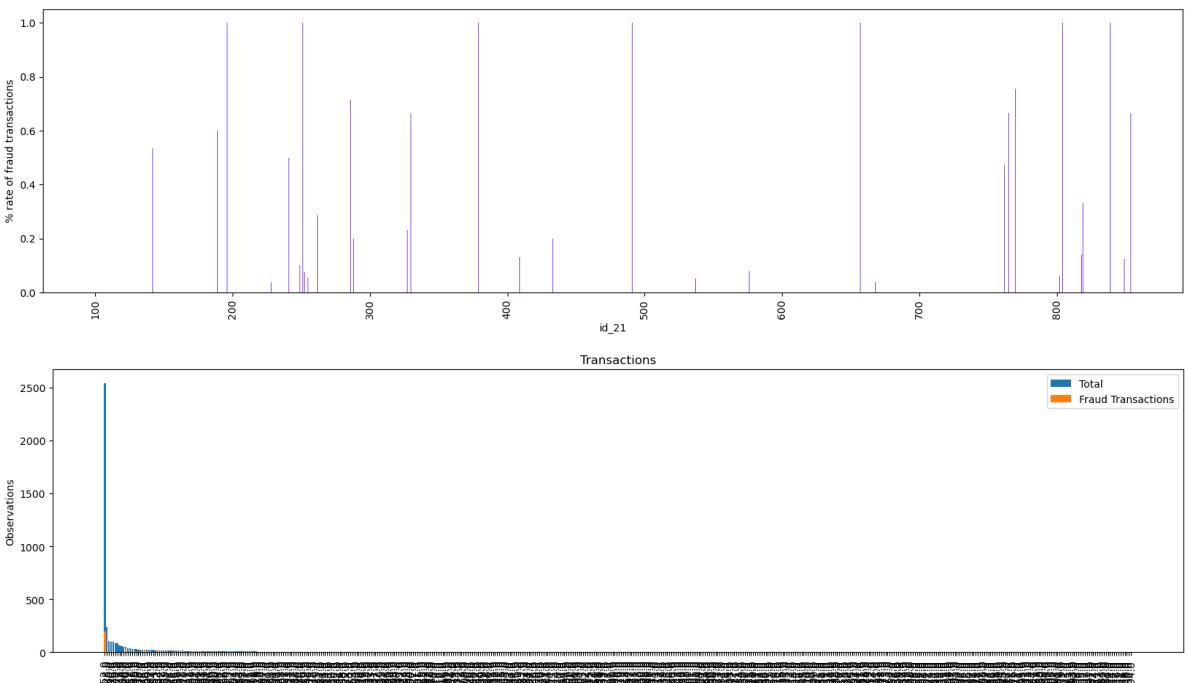
Analysis of id_21 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	5159.000000	count	4733.000000	
count	426.000000			
mean	368.269820	mean	364.47940	
mean	410.382629	std	194.17561	
std	241.117197	min	100.00000	
min	100.000000	25%	252.00000	
min	142.000000	25%	252.00000	
25%	252.000000	50%	252.00000	
50%	252.000000	75%	469.00000	
75%	657.000000	max	854.00000	
max	854.000000	max	854.00000	
Name: id_21, dtype: float64		Name: id_21, dtype: float64		
Name: id_21, dtype: float64				

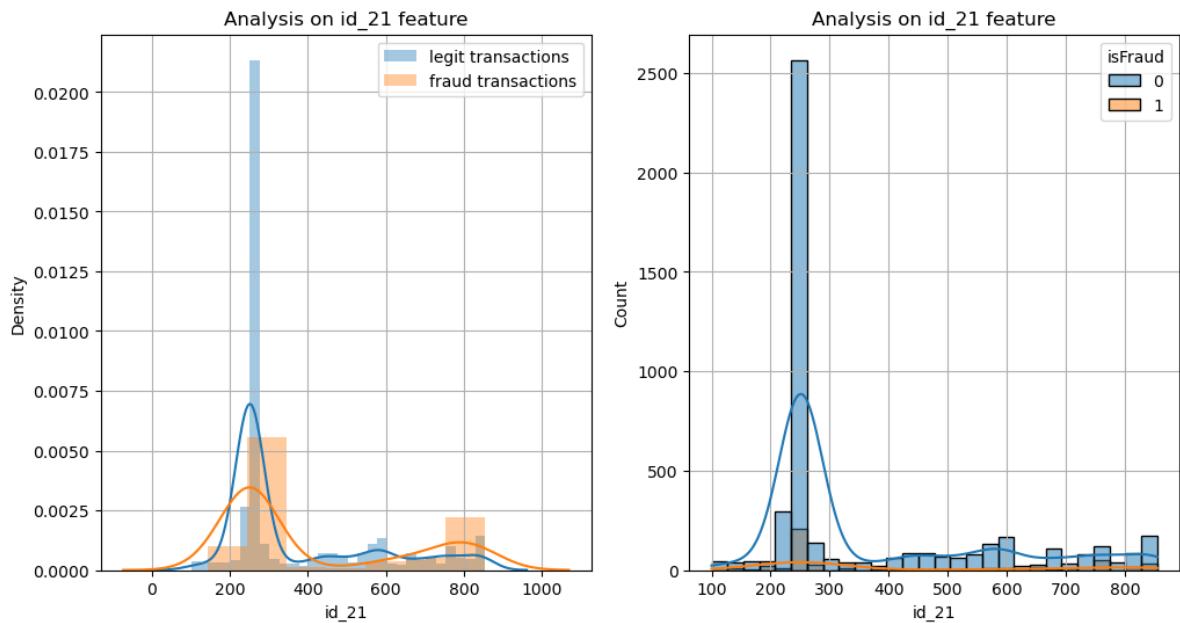
The unique values in id_21 feature is: 490

id_21 feature has 585381 null values, which holds 0.9912639279303688 %



	id_21	isFraud	total	% Occurrence	Avg Fraud
97	252.0	194	2542	0.004305	0.076318
76	228.0	9	239	0.000405	0.037657
100	255.0	6	109	0.000185	0.055046
323	596.0	0	103	0.000174	0.000000
308	576.0	8	101	0.000171	0.079208
..
24	146.0	0	1	0.000002	0.000000
296	558.0	0	1	0.000002	0.000000
297	559.0	0	1	0.000002	0.000000
298	560.0	0	1	0.000002	0.000000
54	194.0	0	1	0.000002	0.000000

[490 rows x 5 columns]



```
*****
*****
```

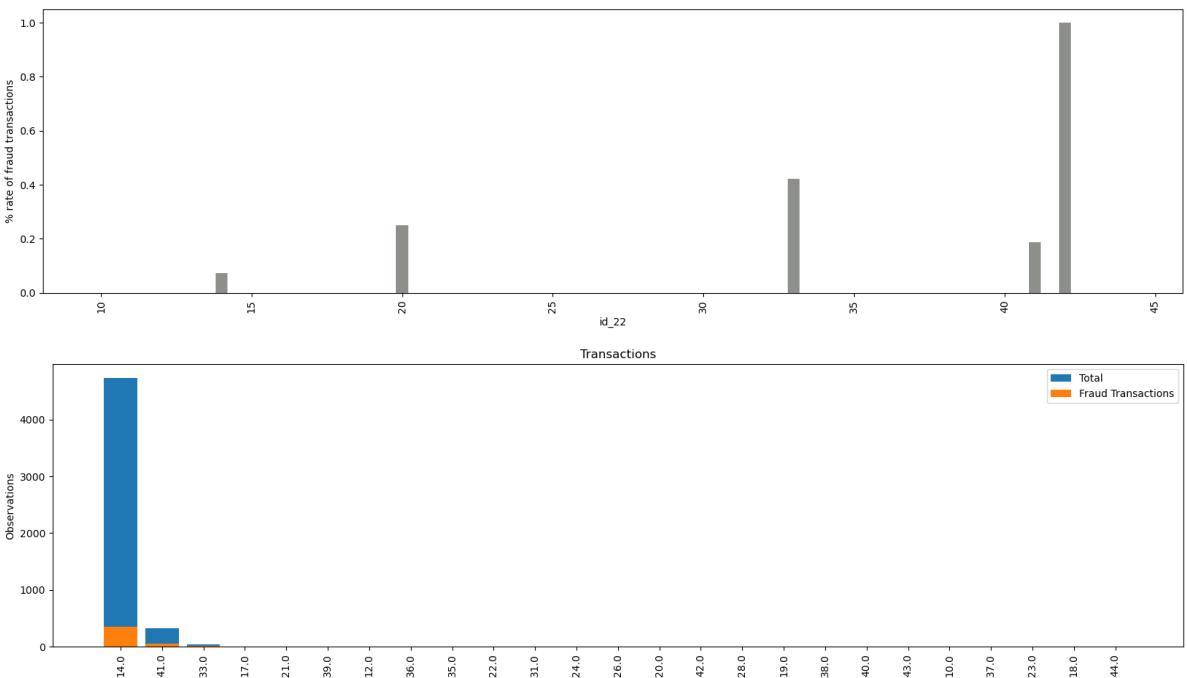
Analysis of id_22 feature

Five Number Summary

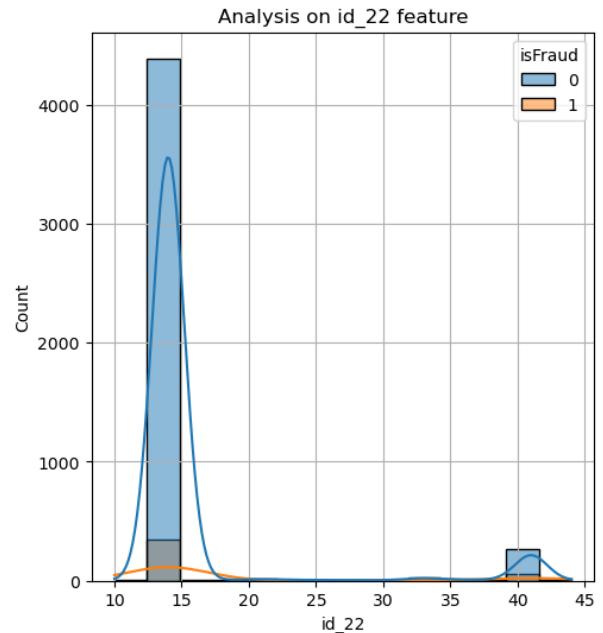
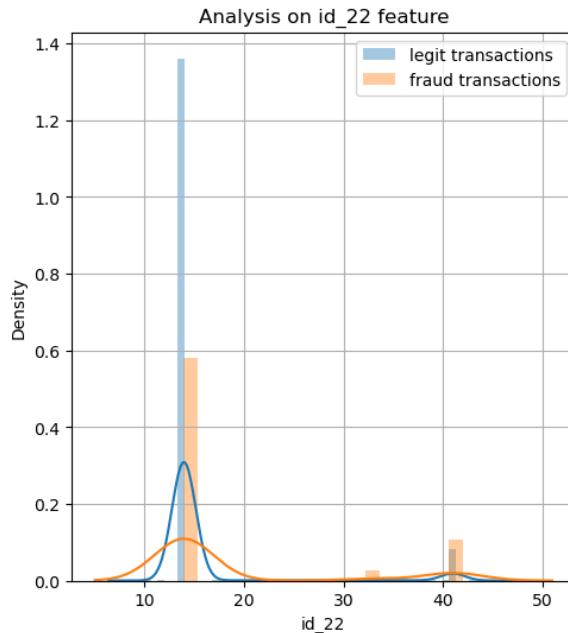
Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	5169.000000	count	4743.000000	
count	426.000000			
mean	16.002708	mean	15.757959	
mean	18.727700	std	6.495174	
std	6.897665	std	9.986157	
std	10.000000	min	10.000000	
min	14.000000	min	14.000000	
25%	14.000000	25%	14.000000	
50%	14.000000	50%	14.000000	
75%	14.000000	75%	14.000000	
max	44.000000	max	44.000000	
max	42.000000			
	Name: id_22, dtype: float64	Name: id_22, dtype: float64		
	Name: id_22, dtype: float64			

The unique values in id_22 feature is: 25

id_22 feature has 585371 null values, which holds 0.9912469942764249 %



	id_22	isFraud	total	% Occurence	Avg Fraud
2	14.0	346	4736	0.008020	0.073057
21	41.0	60	321	0.000544	0.186916
14	33.0	16	38	0.000064	0.421053
3	17.0	0	7	0.000012	0.000000
7	21.0	0	7	0.000012	0.000000
19	39.0	0	6	0.000010	0.000000
1	12.0	0	5	0.000008	0.000000
16	36.0	0	5	0.000008	0.000000
15	35.0	0	5	0.000008	0.000000
8	22.0	0	5	0.000008	0.000000
13	31.0	0	5	0.000008	0.000000
10	24.0	0	4	0.000007	0.000000
11	26.0	0	4	0.000007	0.000000
6	20.0	1	4	0.000007	0.250000
22	42.0	3	3	0.000005	1.000000
12	28.0	0	3	0.000005	0.000000
5	19.0	0	2	0.000003	0.000000
18	38.0	0	2	0.000003	0.000000
20	40.0	0	1	0.000002	0.000000
23	43.0	0	1	0.000002	0.000000
0	10.0	0	1	0.000002	0.000000
17	37.0	0	1	0.000002	0.000000
9	23.0	0	1	0.000002	0.000000
4	18.0	0	1	0.000002	0.000000
24	44.0	0	1	0.000002	0.000000



```
*****
*****
```

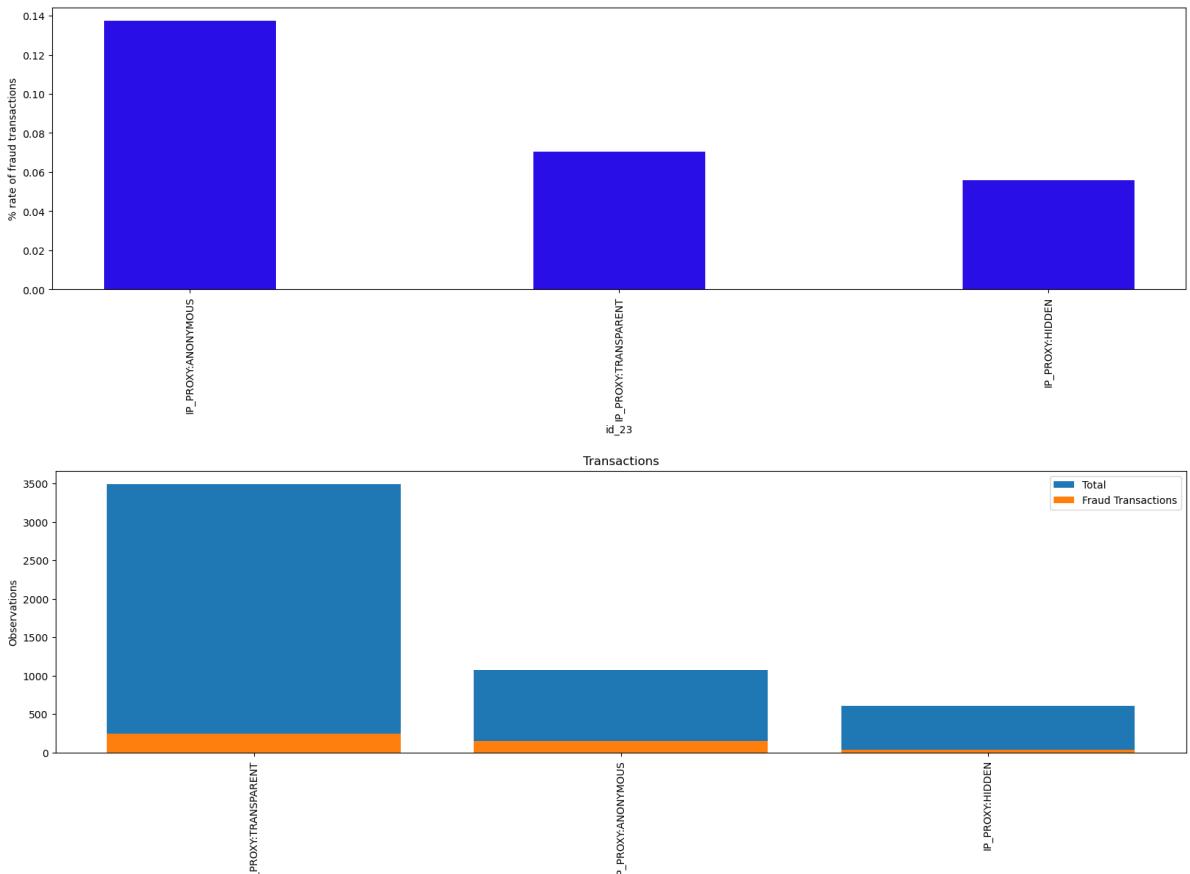
Analysis of id_23 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	5169	count
count	426	
unique	3	unique
unique	3	
top	IP_PROXY:TRANSPARENT	top
top	IP_PROXY:TRANSPARENT	IP_PROXY:TRANSPARENT
freq	3489	freq
freq	245	
Name: id_23, dtype: object		Name: id_23, dtype: object
Name: id_23, dtype: object		

The unique values in id_23 feature is: 3

id_23 feature has 585371 null values, which holds 0.9912469942764249 %



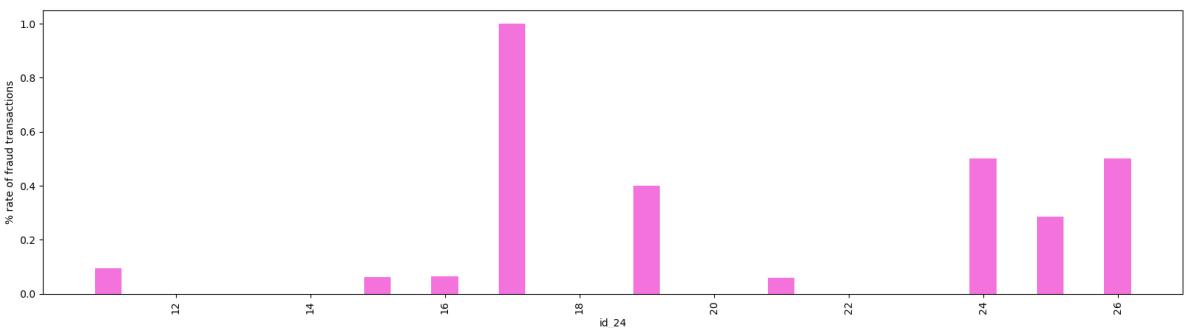
	id_23	isFraud	total	% Occurence	Avg Fraud
2	IP_PROXY:TRANSPARENT	245	3489	0.005908	0.070221
0	IP_PROXY:ANONYMOUS	147	1071	0.001814	0.137255
1	IP_PROXY:HIDDEN	34	609	0.001031	0.055829

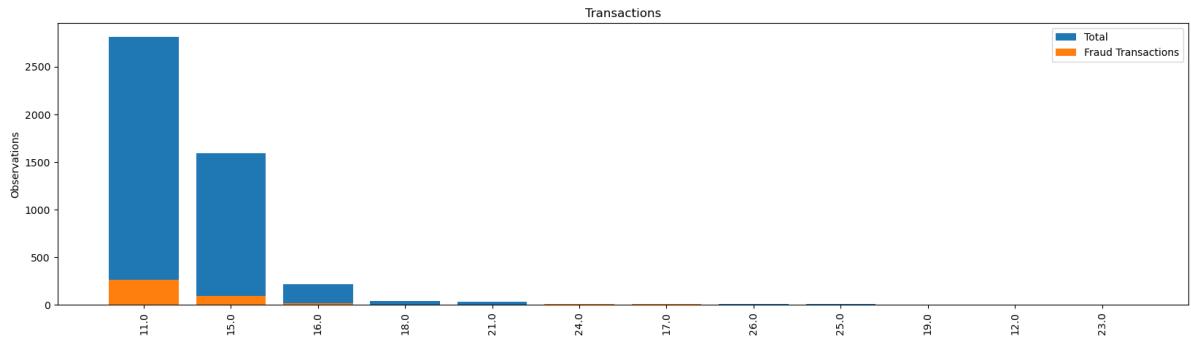
Analysis of id_24 feature

Five Number Summary

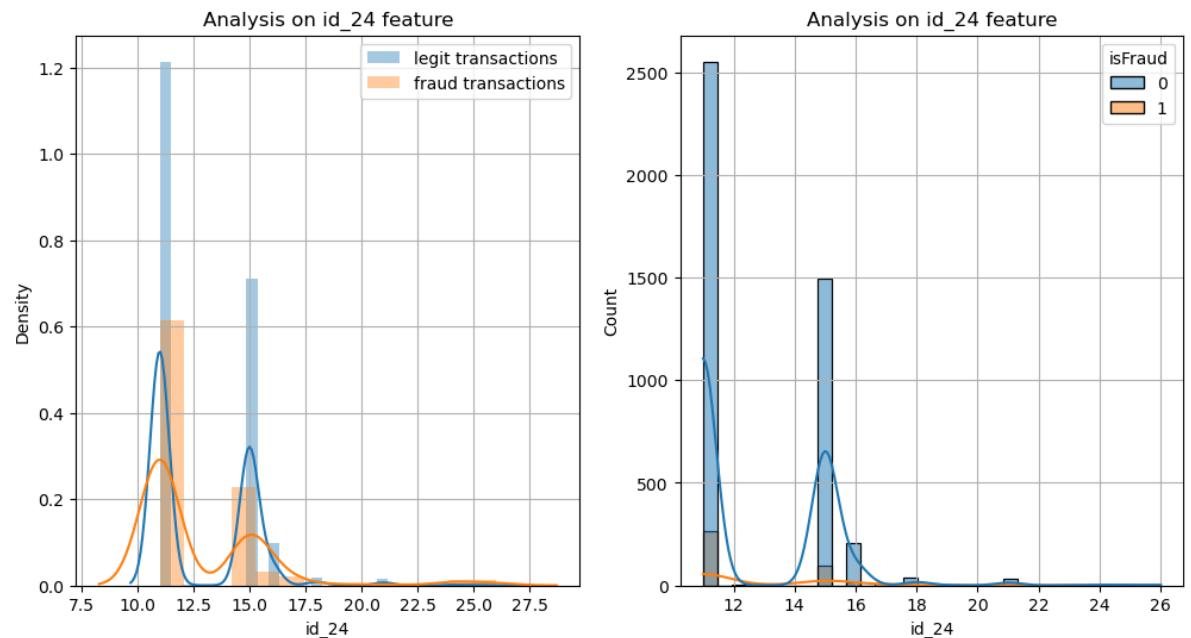
Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	4747.000000	count	4345.000000	
mean	12.800927	mean	12.802301	
std	2.372447	std	2.306966	
min	11.000000	min	11.000000	
25%	11.000000	25%	11.000000	
50%	11.000000	50%	11.000000	
75%	15.000000	75%	15.000000	
max	26.000000	max	26.000000	
Name: id_24, dtype: float64		Name: id_24, dtype: float64		
Name: id_24, dtype: float64				

The unique values in id_24 feature is: 12
 id_24 feature has 585793 null values, which holds 0.9919615944728554 %





	id_24	isFraud	total	% Occurrence	Avg Fraud
0	11.0	265	2817	0.004770	0.094072
2	15.0	98	1594	0.002699	0.061481
3	16.0	14	220	0.000373	0.063636
5	18.0	0	37	0.000063	0.000000
7	21.0	2	33	0.000056	0.060606
9	24.0	6	12	0.000020	0.500000
4	17.0	9	9	0.000015	1.000000
11	26.0	4	8	0.000014	0.500000
10	25.0	2	7	0.000012	0.285714
6	19.0	2	5	0.000008	0.400000
1	12.0	0	4	0.000007	0.000000
8	23.0	0	1	0.000002	0.000000



```
*****
*****
```

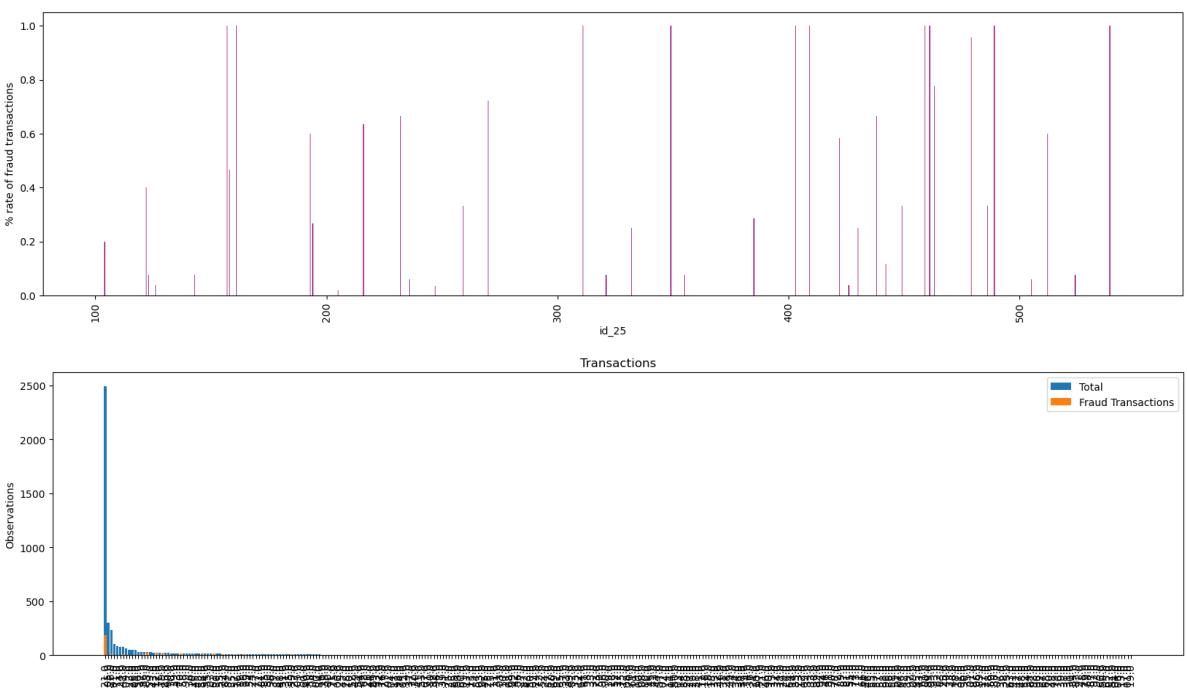
Analysis of id_25 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	5132.000000	count	4715.000000	
count	417.000000			
mean	329.608924	mean	328.622269	
mean	340.764988	std	96.886569	
std	97.461089	std	103.203530	
std	100.000000	min	100.000000	
min	104.000000	min	104.000000	
25%	321.000000	25%	321.000000	
50%	321.000000	50%	321.000000	
75%	371.000000	75%	360.000000	
75%	438.000000	max	371.000000	
max	548.000000	max	548.000000	
	Name: id_25, dtype: float64	Name: id_25, dtype: float64		
	Name: id_25, dtype: float64			

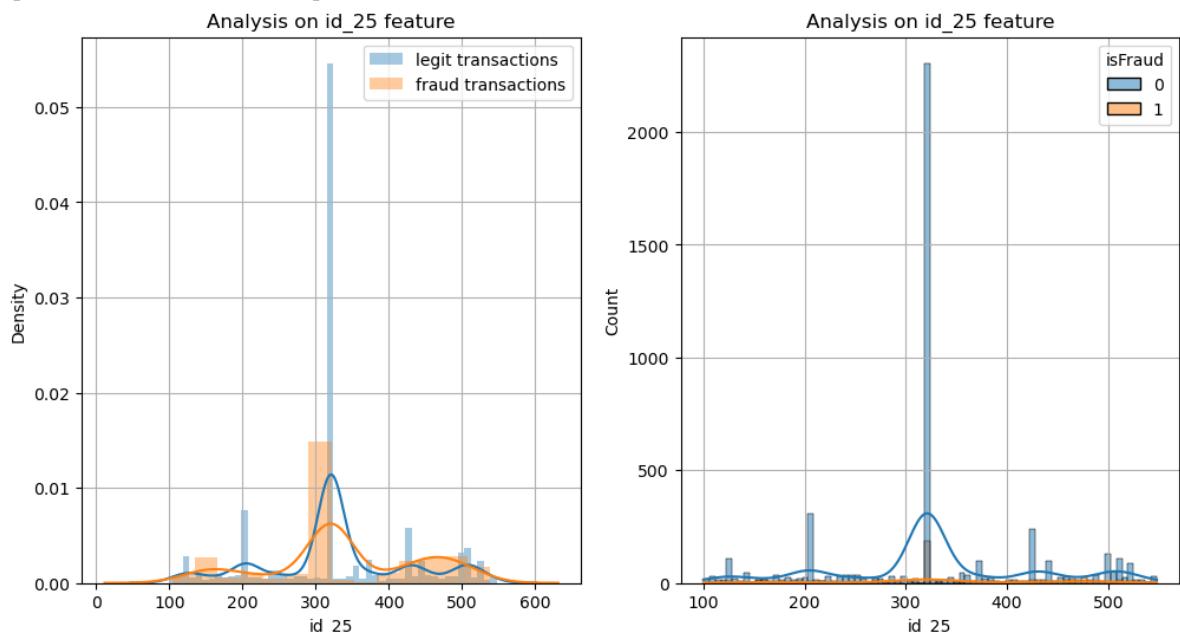
The unique values in id_25 feature is: 341

id_25 feature has 585408 null values, which holds 0.9913096487960172 %



	id_25	isFraud	total	% Occurrence	Avg Fraud
171	321.0	188	2494	0.004223	0.075381
83	205.0	6	301	0.000510	0.019934
250	426.0	9	236	0.000400	0.038136
305	501.0	0	103	0.000174	0.000000
213	371.0	0	83	0.000141	0.000000
..
161	306.0	0	1	0.000002	0.000000
168	313.0	0	1	0.000002	0.000000
169	317.0	0	1	0.000002	0.000000
1	101.0	0	1	0.000002	0.000000
170	319.0	0	1	0.000002	0.000000

[341 rows x 5 columns]



```
*****
*****
```

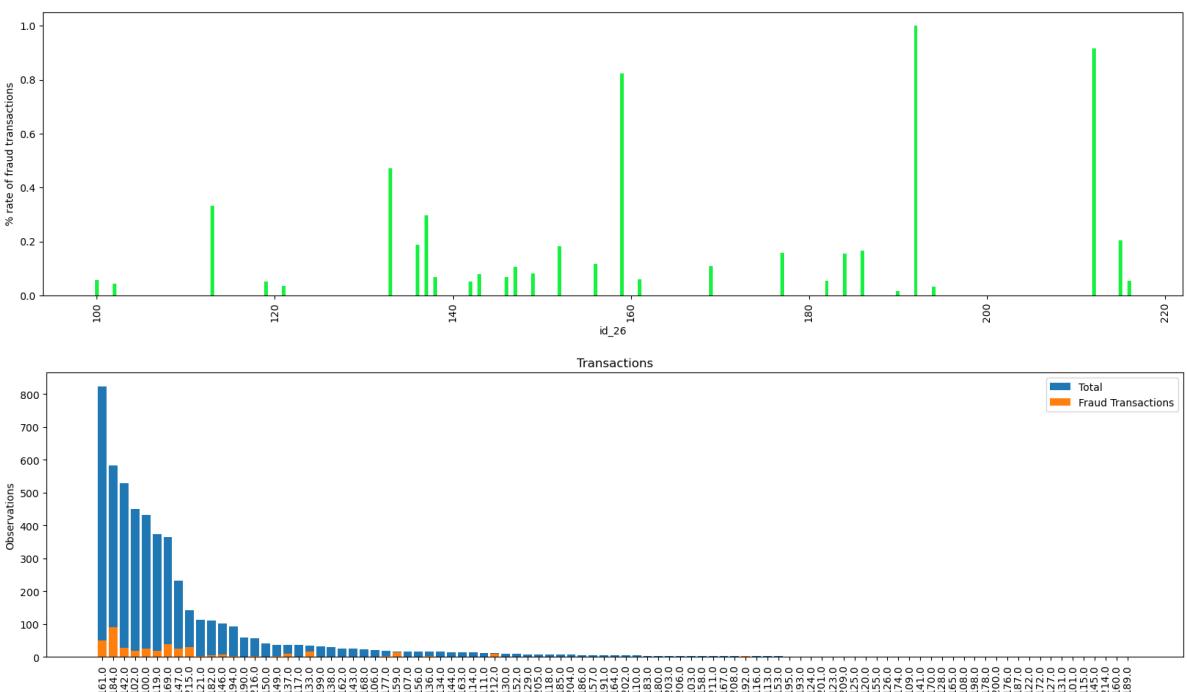
Analysis of id_26 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	5163.000000	count	4738.000000	
count	425.000000			
mean	149.070308	mean	148.112917	
mean	159.743529	std	31.961300	
std	32.101995	std	31.768511	
std	31.768511	min	100.000000	
min	100.000000	min	100.000000	
min	100.000000	25%	119.000000	
25%	119.000000	25%	119.000000	
50%	142.000000	50%	147.000000	
50%	147.000000	75%	161.000000	
75%	161.000000	75%	169.000000	
75%	169.000000	max	184.000000	
max	184.000000	max	216.000000	
max	216.000000	max	216.000000	
	Name: id_26, dtype: float64	Name: id_26, dtype: float64		
	Name: id_26, dtype: float64			

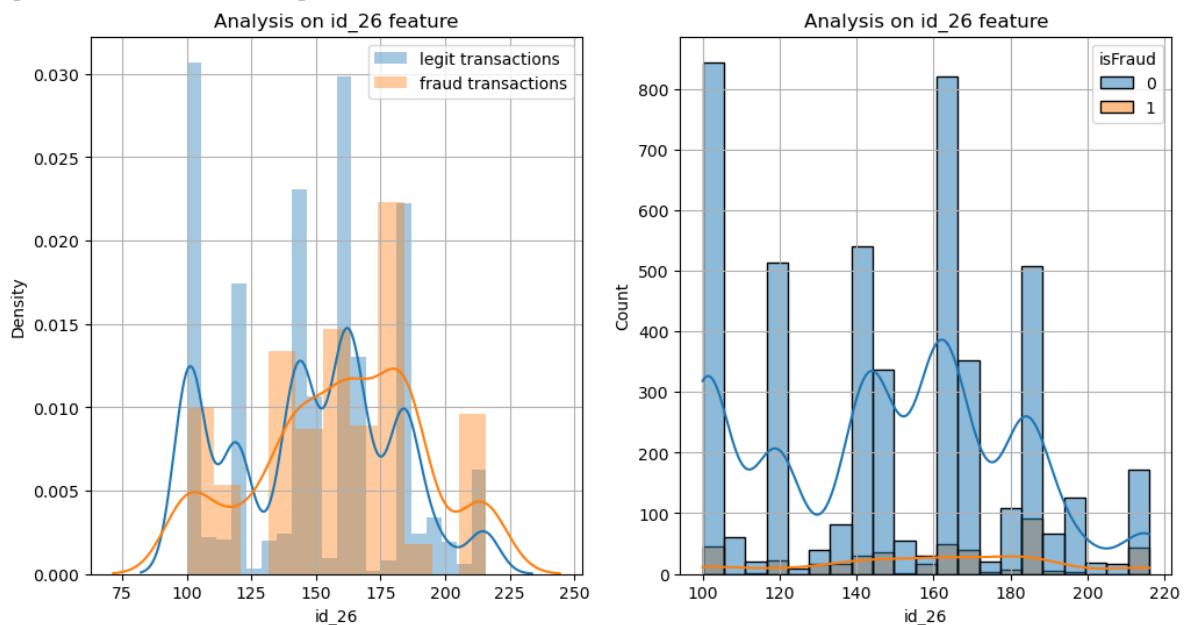
The unique values in id_26 feature is: 95

id_26 feature has 585377 null values, which holds 0.9912571544687913 %



	id_26	isFraud	total	% Occurence	Avg Fraud
51	161.0	50	824	0.001395	0.060680
68	184.0	91	582	0.000986	0.156357
35	142.0	28	528	0.000894	0.053030
2	102.0	20	451	0.000764	0.044346
0	100.0	25	433	0.000733	0.057737
..
12	115.0	0	1	0.000002	0.000000
38	145.0	0	1	0.000002	0.000000
92	214.0	0	1	0.000002	0.000000
50	160.0	0	1	0.000002	0.000000
72	189.0	0	1	0.000002	0.000000

[95 rows x 5 columns]



```
*****
*****
```

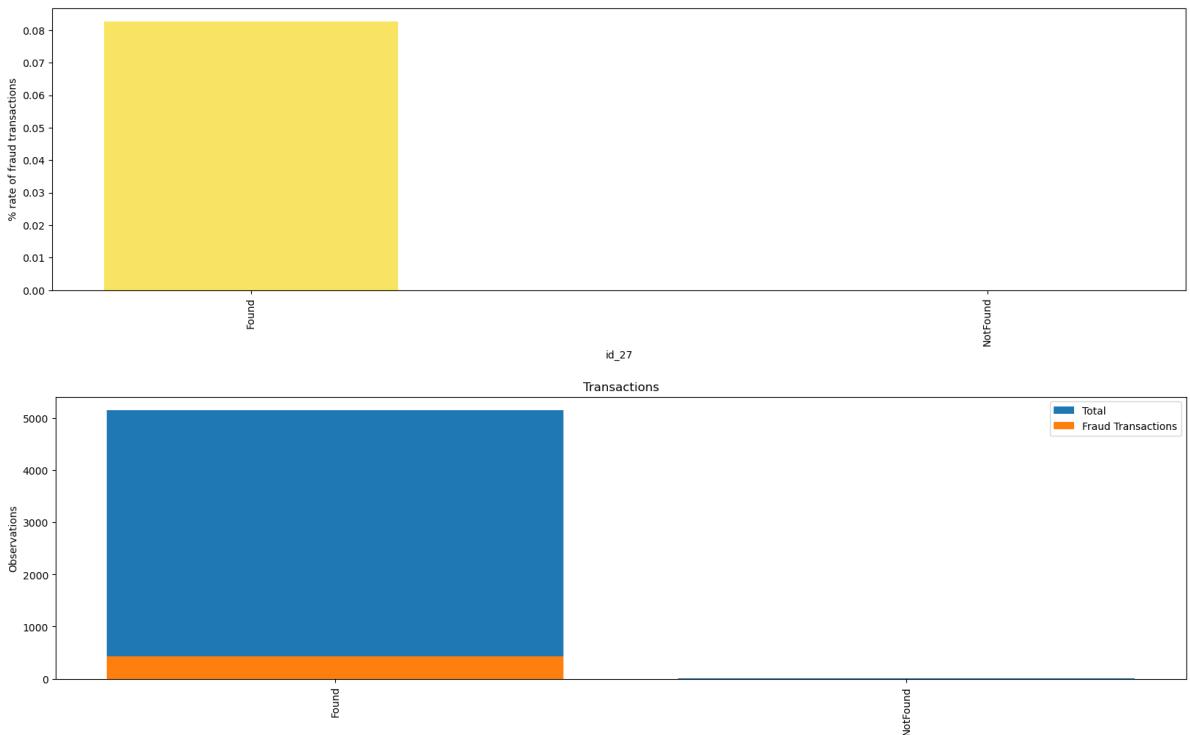
Analysis of id_27 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	5169	count	4743	
unique	2	unique	2	
freq	5155	freq	4729	
Name: id_27, dtype: object		Name: id_27, dtype: object		Name: id_27, dtype: object
Name: id_27, dtype: object				

The unique values in id_27 feature is: 2

id_27 feature has 585371 null values, which holds 0.9912469942764249 %



	id_27	isFraud	total	% Occurence	Avg Fraud
0	Found	426	5155	0.008729	0.082638
1	NotFound	0	14	0.000024	0.000000

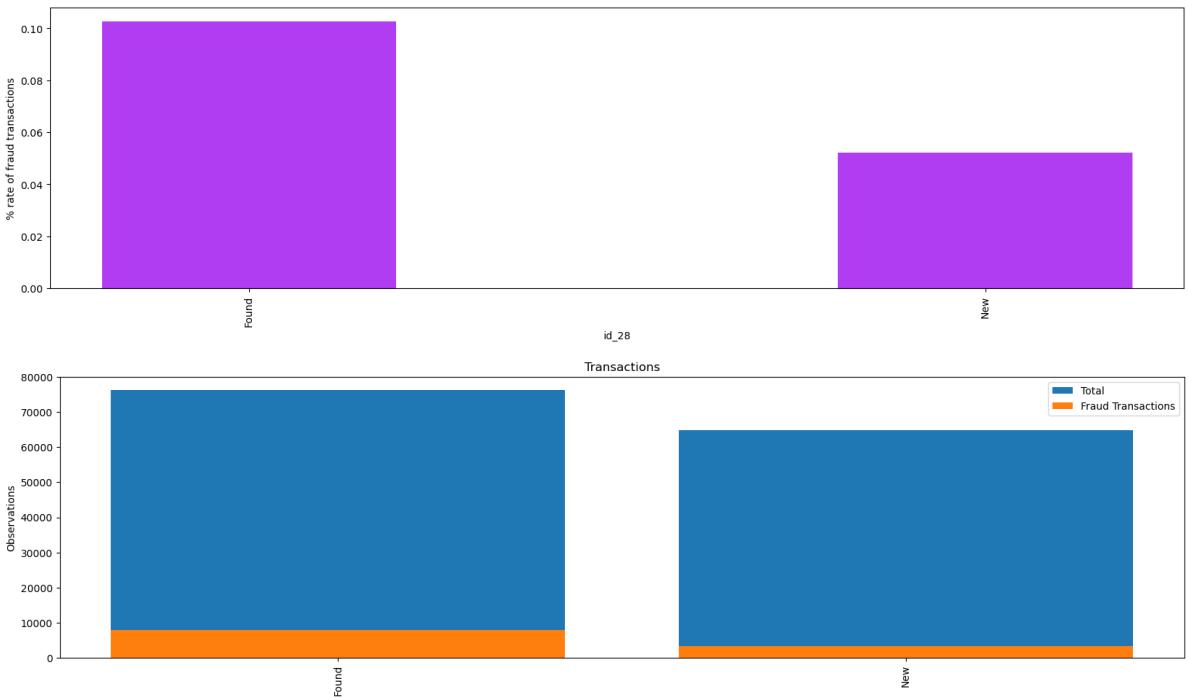
Analysis of id_28 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	140978	count	129760	count	11218
unique	2	unique	2	unique	2
top	Found	top	Found	top	Found
freq	76232	freq	68399	freq	7833
Name:	id_28, dtype: object	Name:	id_28, dtype: object	Name:	id_28, dtype: object
Name:	id_28, dtype: object				

The unique values in id_28 feature is: 2

id_28 feature has 449562 null values, which holds 0.7612727334304196 %



	id_28	isFraud	total	% Occurrence	Avg Fraud
0	Found	7833	76232	0.129089	0.102752
1	New	3385	64746	0.109639	0.052281

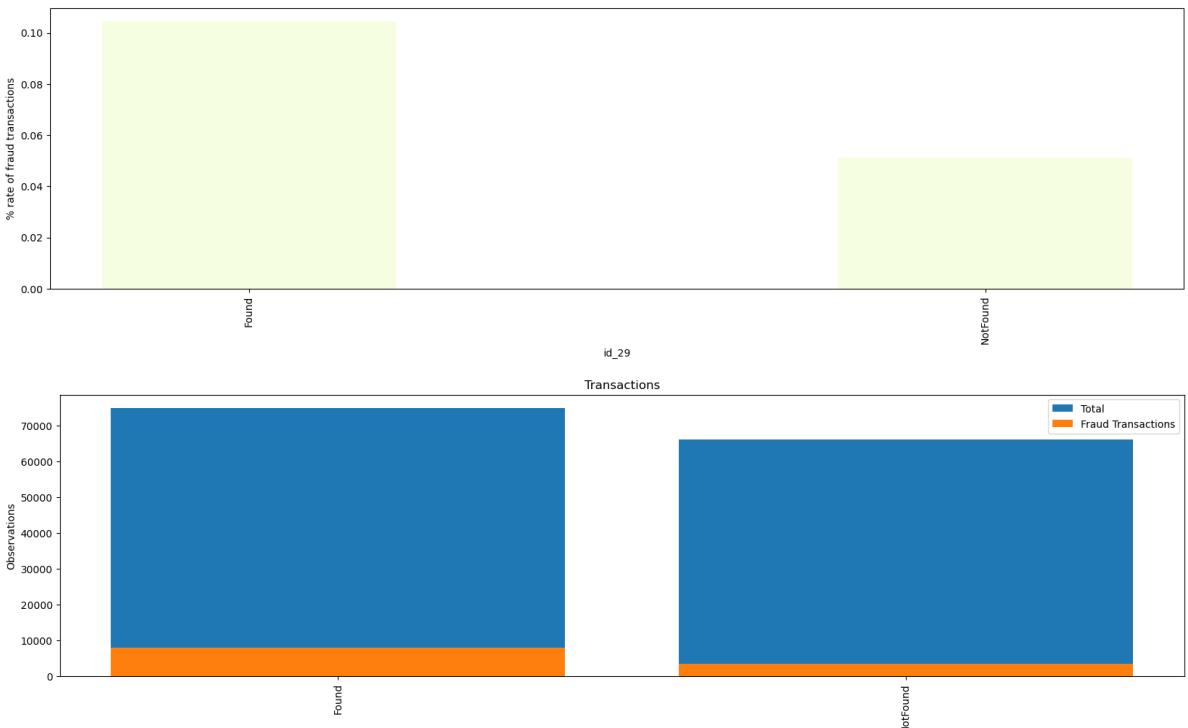
Analysis of id_29 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	140978	count	129760	count	11218
unique	2	unique	2	unique	2
top	Found	top	Found	top	Found
freq	74926	freq	67099	freq	7827
Name:	id_29, dtype: object	Name:	id_29, dtype: object	Name:	id_29, dtype: object

The unique values in id_29 feature is: 2

id_29 feature has 449562 null values, which holds 0.7612727334304196 %



	id_29	isFraud	total	% Occurence	Avg Fraud
0	Found	7827	74926	0.126877	0.104463
1	NotFound	3391	66052	0.111850	0.051338

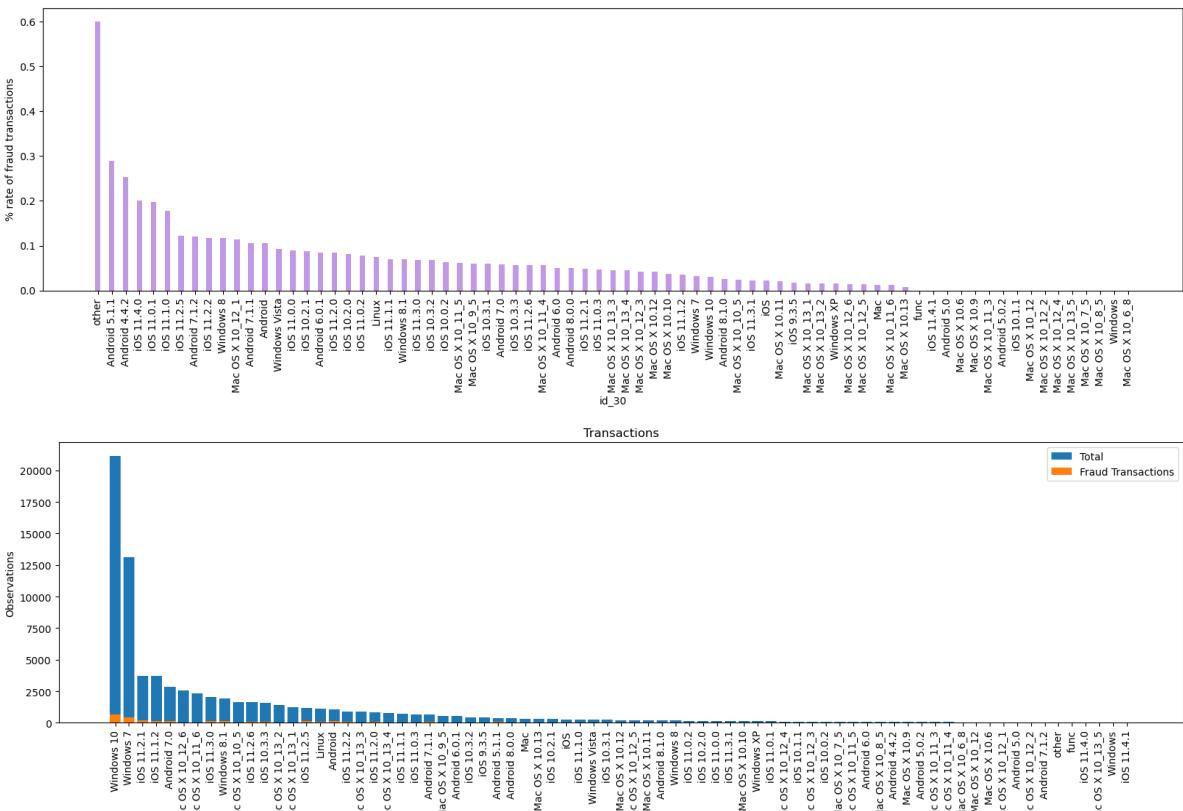
Analysis of id_30 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count		77565		
count	3442			
unique	75			
unique	59			
top	Windows 10			
top	Windows 10			
freq	21155			
freq	657			
Name: id_30, dtype: object		Name: id_30, dtype: object		
Name: id_30, dtype: object				

The unique values in id_30 feature is: 75

id_30 feature has 512975 null values, which holds 0.868654113184543 %



		id_30	isFraud	total	% Occurence	Avg Fraud
42		Windows 10	657	21155	0.035823	0.031056
43		Windows 7	432	13110	0.022200	0.032952
65		iOS 11.2.1	183	3722	0.006303	0.049167
63		iOS 11.1.2	130	3699	0.006264	0.035145
7		Android 7.0	166	2871	0.004862	0.057820
..	
48		func	0	10	0.000017	0.000000
71		iOS 11.4.0	1	5	0.000008	0.200000
36	Mac OS X 10_13_5		0	4	0.000007	0.000000
41		Windows	0	3	0.000005	0.000000
72		iOS 11.4.1	0	1	0.000002	0.000000

[75 rows x 5 columns]

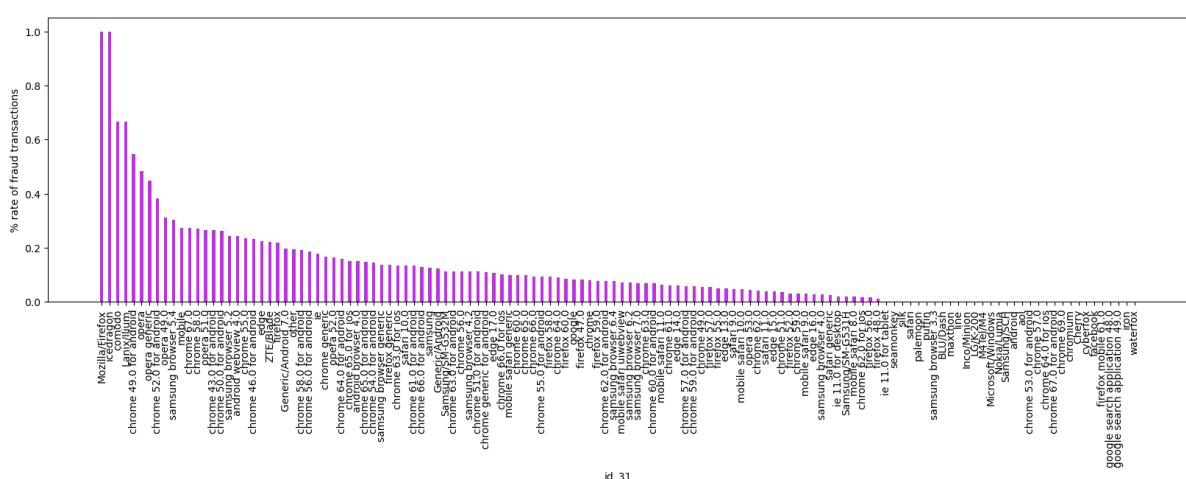
Analysis of id 31 feature

Five Number Summary

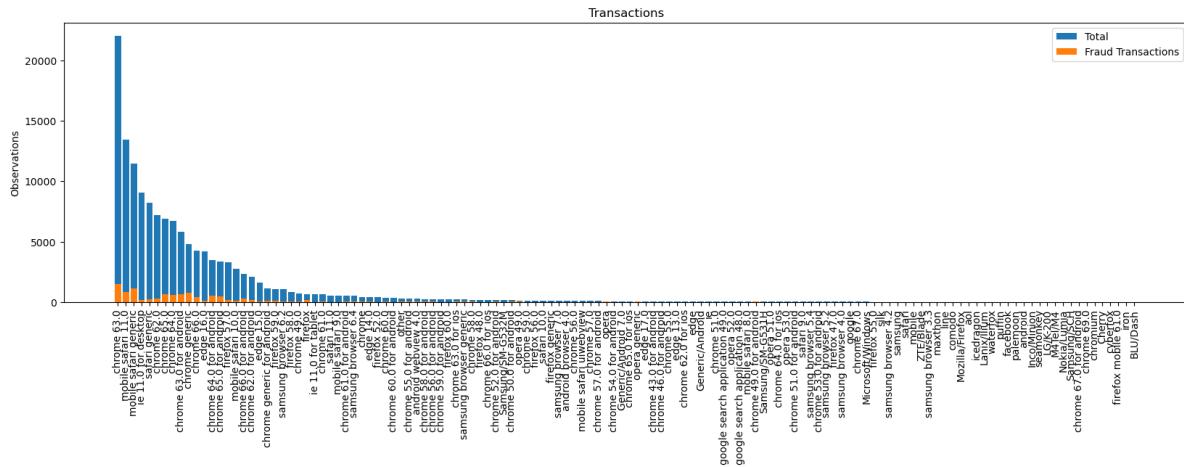
Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	140282	count	129111	
unique	11171	unique	128	
top	chrome 63.0	top	chrome 63.0	
freq	22000	freq	20497	
		Name: id_31, dtype: object	Name: id_31, dtype: object	
		Name: id_31, dtype: object		

The unique values in id 31 feature is: 130

id_31 feature has 450258 null values, which holds 0.7624513157449114 %



1. self_case_study_1_IEEE_fraud_detection_EDA



		id_31	isFraud	total	% Occurrence	Avg Fraud
47	chrome	63.0	1503	22000	0.037254	0.068318
98	mobile safari	11.0	842	13423	0.022730	0.062728
101	mobile safari generic		1146	11474	0.019430	0.099878
91	ie 11.0 for desktop		175	9030	0.015291	0.019380
116	safari generic		205	8195	0.013877	0.025015
..
1	Cherry		0	1	0.000002	0.000000
66	cyberfox		0	1	0.000002	0.000000
85	firefox mobile	61.0	0	1	0.000002	0.000000
93	iron		0	1	0.000002	0.000000
0	BLU/Dash		0	1	0.000002	0.000000

[130 rows x 5 columns]

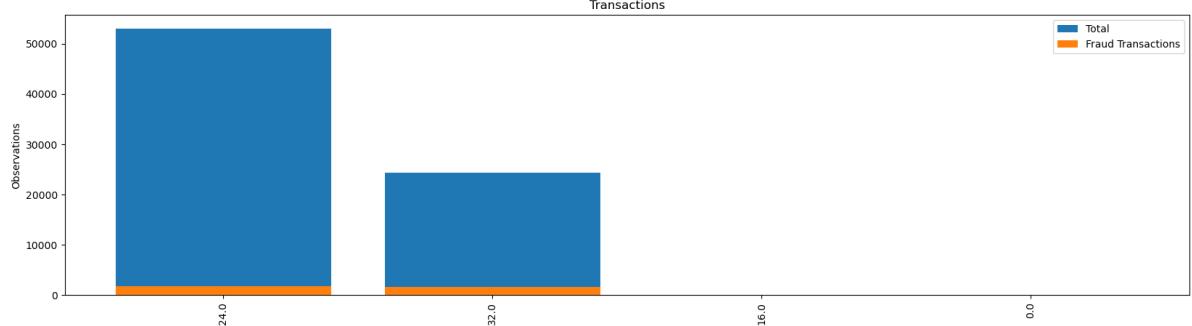
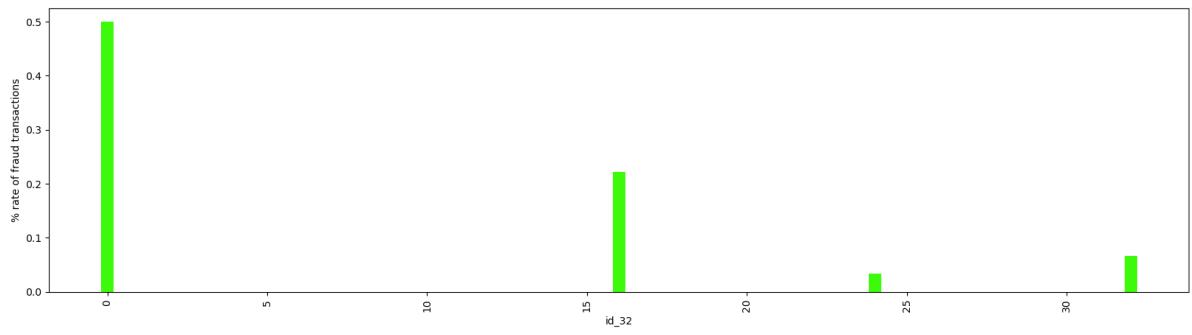
Analysis of id_32 feature

Five Number Summary

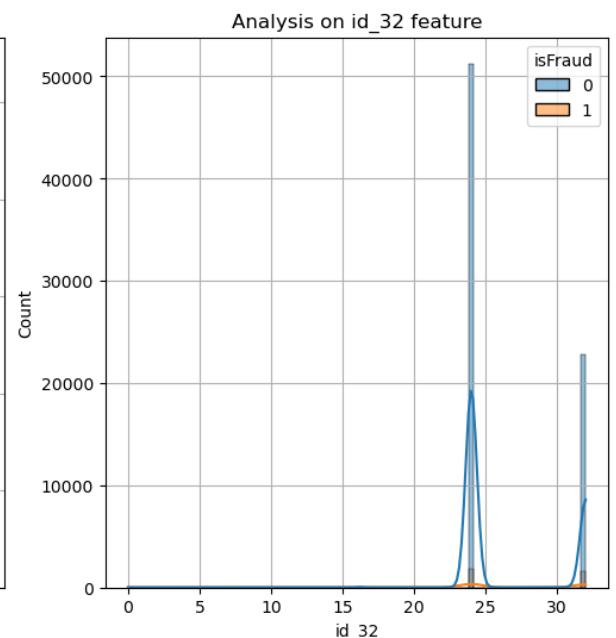
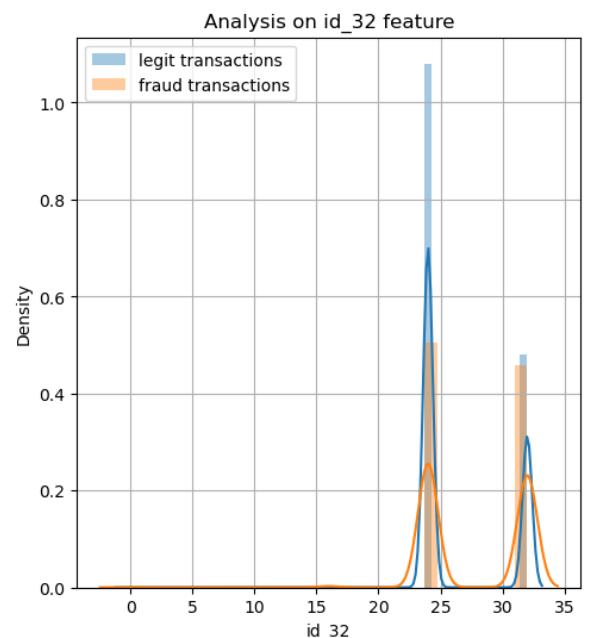
Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	77586.000000	count	74135.000000	
mean	3451.000000	mean	26.452391	
std	26.508597	std	27.716024	
min	3.737502	min	4.154470	
25%	0.000000	25%	0.000000	
50%	24.000000	50%	24.000000	
75%	24.000000	75%	24.000000	
max	32.000000	max	32.000000	
	Name: id_32, dtype: float64		Name: id_32, dtype: float64	
	Name: id_32, dtype: float64			

The unique values in id_32 feature is: 4

id_32 feature has 512954 null values, which holds 0.8686185525112609 %



	id_32	isFraud	total	% Occurrence	Avg Fraud
2	24.0	1	1800	0.089869	0.033917
3	32.0	1	1630	0.041366	0.066727
1	16.0	0	18	0.000137	0.222222
0	0.0	0	3	0.000010	0.500000



```
*****
*****
```

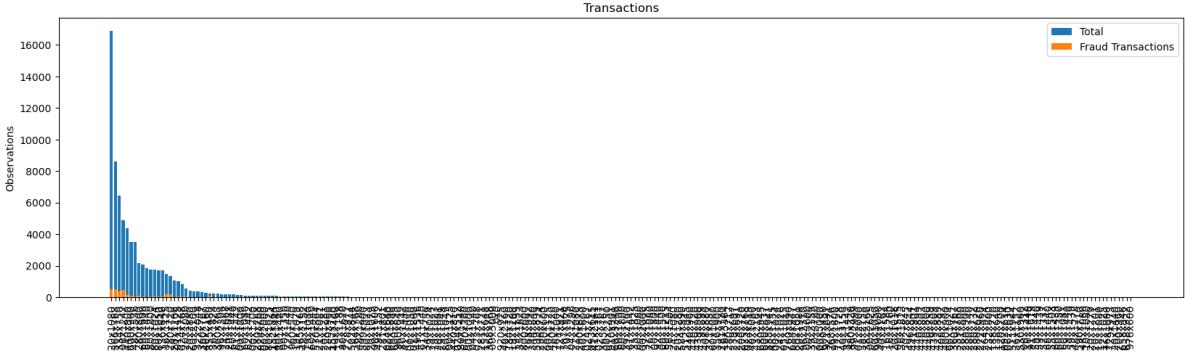
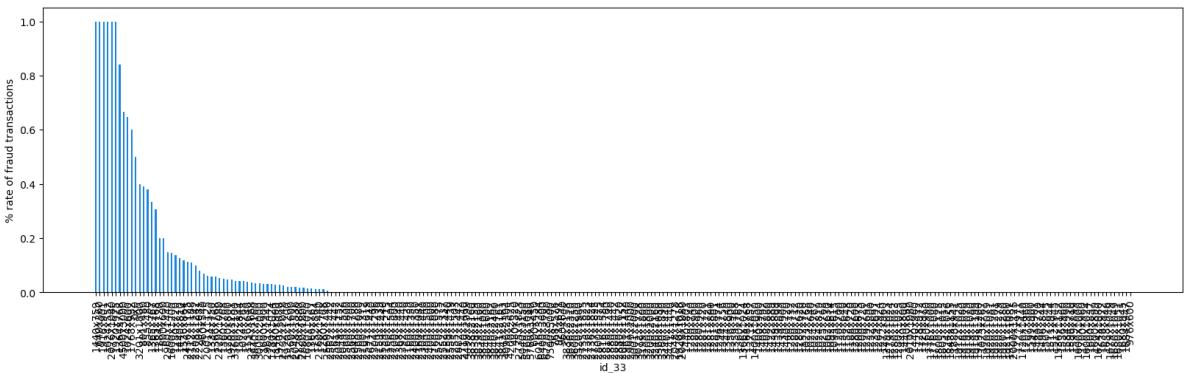
Analysis of id_33 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	73289	count	69926	
count	3363	unique	254	
unique	260	top	1920x1080	
unique	59	freq	16384	
top	1920x1080	Name: id_33, dtype: object	Name: id_33, dtype: object	
top	1366x768	Name: id_33, dtype: object		
freq	16874			
freq	493			
Name: id_33, dtype: object				
Name: id_33, dtype: object				

The unique values in id_33 feature is: 260

id_33 feature has 517251 null values, which holds 0.8758949436109323 %



	id_33	isFraud	total	% Occurrence	Avg Fraud
132	1920x1080	490	16874	0.028574	0.029039
60	1366x768	493	8605	0.014571	0.057292
48	1334x750	381	6447	0.010917	0.059097
163	2208x1242	482	4900	0.008297	0.098367
77	1440x900	128	4384	0.007424	0.029197
..
110	1720x1440	0	1	0.000002	0.000000
108	1700x960	0	1	0.000002	0.000000
101	1658x946	0	1	0.000002	0.000000
100	1638x922	0	1	0.000002	0.000000
259	976x600	0	1	0.000002	0.000000

[260 rows x 5 columns]

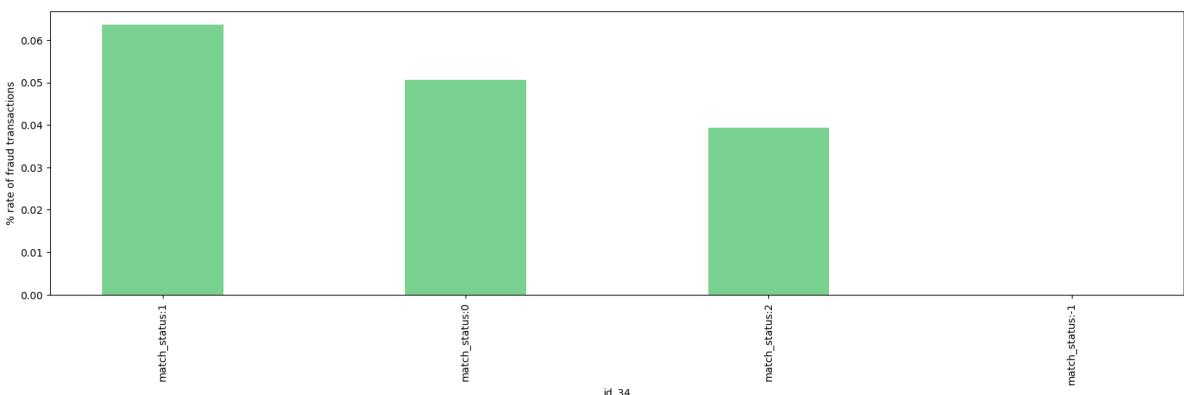
Analysis of id_34 feature

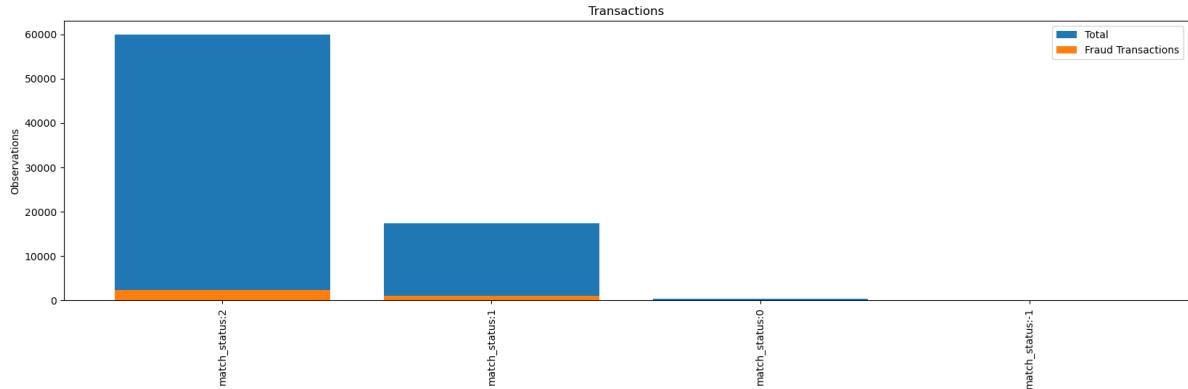
Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	77805	count	74320	count	3485
unique	4	unique	4	unique	3
top	match_status:2	top	match_status:2	top	match_status:2
freq	60011	freq	57653	freq	2358
Name: id_34, dtype: object		Name: id_34, dtype: object		Name: id_34, dtype: object	
Name: id_34, dtype: object					

The unique values in id_34 feature is: 4

id_34 feature has 512735 null values, which holds 0.8682477054898906 %





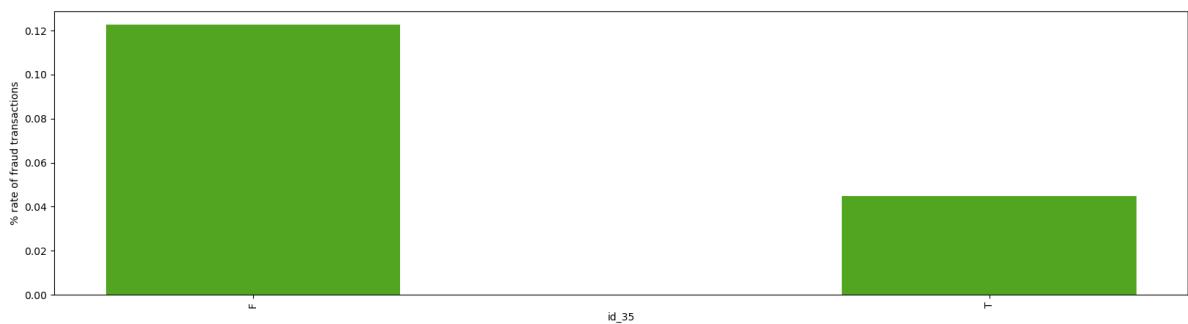
	id_34	isFraud	total	% Occurence	Avg Fraud
3	match_status:2	2358	60011	0.101621	0.039293
2	match_status:1	1106	17376	0.029424	0.063651
1	match_status:0	21	415	0.000703	0.050602
0	match_status:-1	0	3	0.000005	0.000000

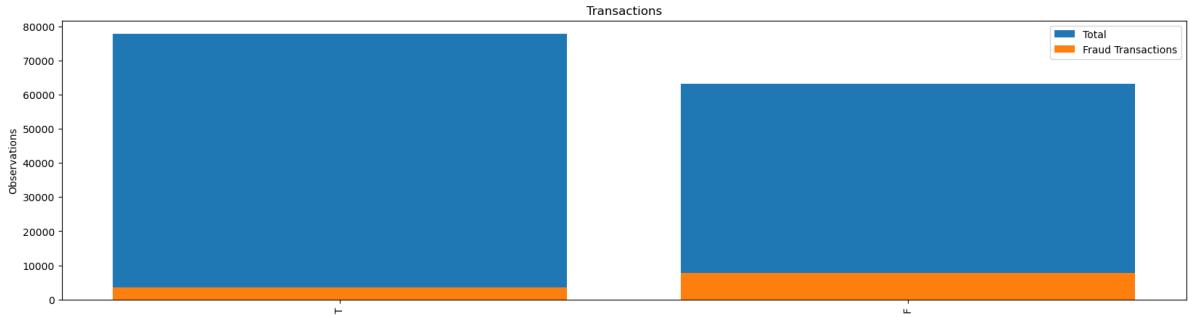
Analysis of id_35 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	140985	count	129763	
unique	2	unique	2	
top	T	top	T	
freq	77814	freq	74337	
	Name: id_35, dtype: object		Name: id_35, dtype: object	
	Name: id_35, dtype: object			

The unique values in id_35 feature is: 2
 id_35 feature has 449555 null values, which holds 0.7612608798726589 %





	id_35	isFraud	total	% Occurrence	Avg Fraud
1	T	3477	77814	0.131768	0.044683
0	F	7745	63171	0.106972	0.122604

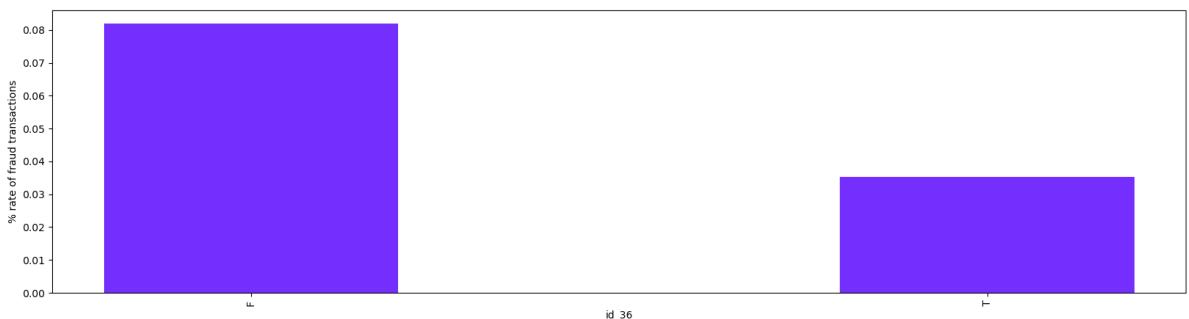
Analysis of id_36 feature

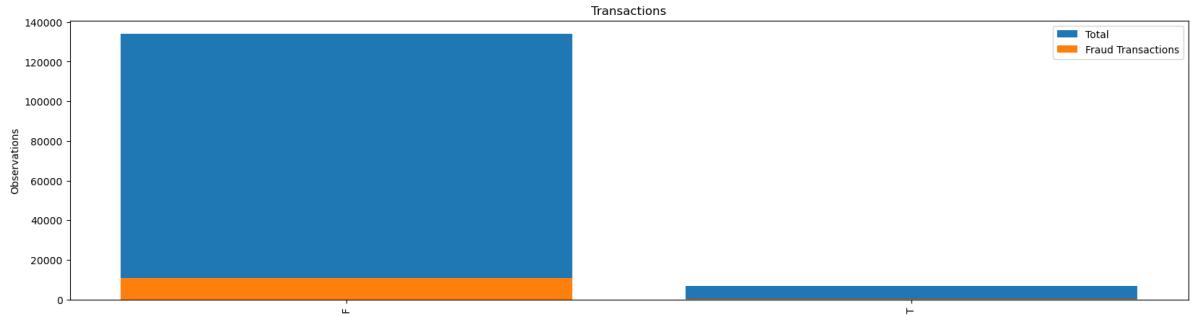
Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	140985	count	129763	
count	11222	unique	2	
unique	2	top	F	
top	F	freq	134066	
freq	10978	freq	123088	
Name: id_36, dtype: object		Name: id_36, dtype: object		
Name: id_36, dtype: object				

The unique values in id_36 feature is: 2

id_36 feature has 449555 null values, which holds 0.7612608798726589 %





id_36	isFraud	total	% Occurrence	Avg Fraud
0	F	10978	0.227023	0.081885
1	T	244	0.011716	0.035265

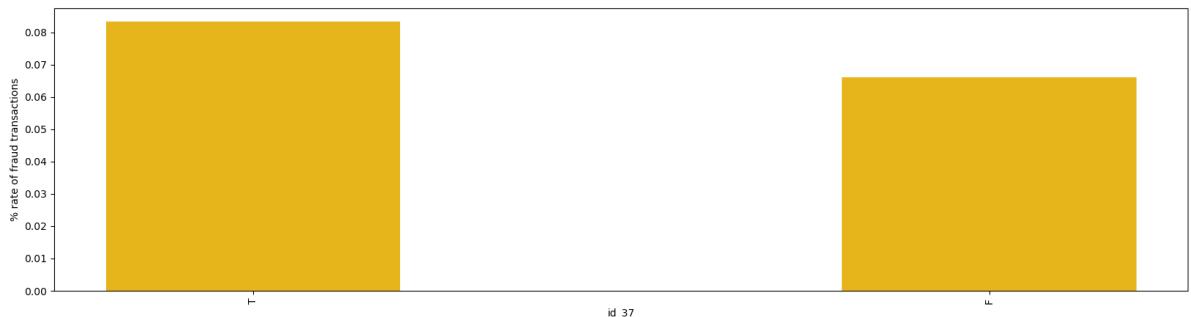
Analysis of id_37 feature

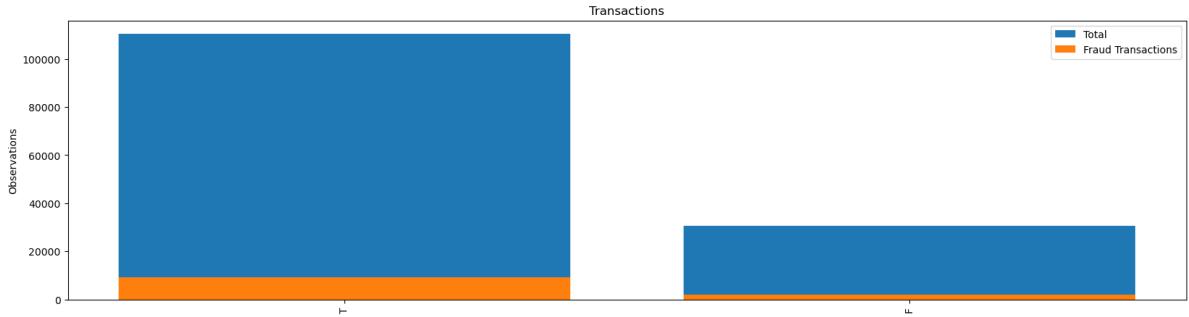
Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	140985	count	129763	
unique	11222	unique	2	
top	2	top	T	
freq	110452	freq	101250	
Name: id_37, dtype: object		Name: id_37, dtype: object		
Name: id_37, dtype: object				

The unique values in id_37 feature is: 2

id_37 feature has 449555 null values, which holds 0.7612608798726589 %





	id_37	isFraud	total	% Occurrence	Avg Fraud
1	T	9202	110452	0.187036	0.083312
0	F	2020	30533	0.051704	0.066158

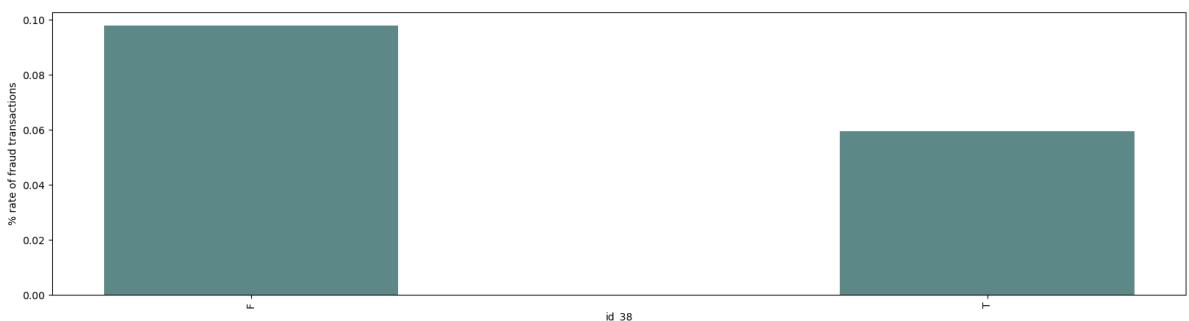
Analysis of id_38 feature

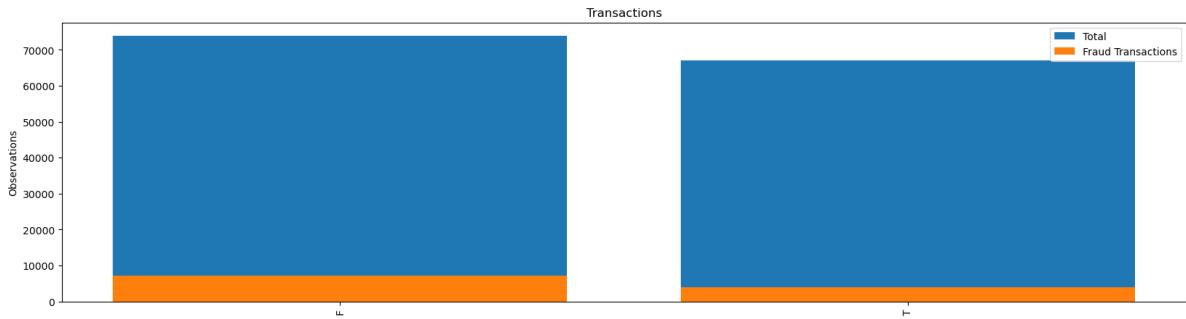
Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	140985	count	129763	count	11222
unique	2	unique	2	unique	2
top	F	top	F	top	F
freq	73922	freq	66693	freq	7229
Name:	id_38, dtype: object	Name:	id_38, dtype: object	Name:	id_38, dtype: object
Name:	id_38, dtype: object				

The unique values in id_38 feature is: 2

id_38 feature has 449555 null values, which holds 0.7612608798726589 %





	id_38	isFraud	total	% Occurrence	Avg Fraud
0	F	7229	73922	0.125177	0.097792
1	T	3993	67063	0.113562	0.059541

Observation:

1. While some of the id_features are encoded, others are self-explanatory.
2. id_28 has values as "Found" and "New", this feature may correspond to the history of the transaction i.e. whether any transaction with the same details has been done before or this is the first time this transaction is happening.
3. The id_30 characteristic is related to the operating system utilised by the transaction device.
4. The browser being used to complete the transaction is identified by the id_31.
5. The dimensions of the device being used for the transaction are represented by the id_33.

1.2 Univariate Analysis of numerical features

```
In [57]: # available numerical features
num_fea = ['TransactionDT', 'TransactionAmt', 'dist1', 'dist2', 'C1', 'C2', 'C3',
'C11', 'C12', 'C13', 'C14', 'D1', 'D2', 'D3', 'D4', 'D5', 'D6', 'D7',
'D12', 'D13', 'D14', 'D15', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7',
'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20', 'V21',
'V26', 'V27', 'V28', 'V29', 'V30', 'V31', 'V32', 'V33', 'V34', 'V35',
'V40', 'V41', 'V42', 'V43', 'V44', 'V45', 'V46', 'V47', 'V48', 'V49',
'V54', 'V55', 'V56', 'V57', 'V58', 'V59', 'V60', 'V61', 'V62', 'V63',
'V68', 'V69', 'V70', 'V71', 'V72', 'V73', 'V74', 'V75', 'V76', 'V77',
'V82', 'V83', 'V84', 'V85', 'V86', 'V87', 'V88', 'V89', 'V90', 'V91',
'V96', 'V97', 'V98', 'V99', 'V100', 'V101', 'V102', 'V103', 'V104', 'V105',
'V109', 'V110', 'V111', 'V112', 'V113', 'V114', 'V115', 'V116', 'V117',
'V121', 'V122', 'V123', 'V124', 'V125', 'V126', 'V127', 'V128', 'V129',
'V133', 'V134', 'V135', 'V136', 'V137', 'V138', 'V139', 'V140', 'V141',
'V145', 'V146', 'V147', 'V148', 'V149', 'V150', 'V151', 'V152', 'V153',
'V157', 'V158', 'V159', 'V160', 'V161', 'V162', 'V163', 'V164', 'V165',
'V169', 'V170', 'V171', 'V172', 'V173', 'V174', 'V175', 'V176', 'V177',
'V181', 'V182', 'V183', 'V184', 'V185', 'V186', 'V187', 'V188', 'V189',
'V193', 'V194', 'V195', 'V196', 'V197', 'V198', 'V199', 'V200', 'V201',
'V205', 'V206', 'V207', 'V208', 'V209', 'V210', 'V211', 'V212', 'V213']
```

1. self_case_study_1_IEEE_fraud_detection_EDA

```
'V217', 'V218', 'V219', 'V220', 'V221', 'V222', 'V223', 'V224', 'V225'
'V229', 'V230', 'V231', 'V232', 'V233', 'V234', 'V235', 'V236', 'V237'
'V241', 'V242', 'V243', 'V244', 'V245', 'V246', 'V247', 'V248', 'V249'
'V253', 'V254', 'V255', 'V256', 'V257', 'V258', 'V259', 'V260', 'V261'
'V265', 'V266', 'V267', 'V268', 'V269', 'V270', 'V271', 'V272', 'V273'
'V277', 'V278', 'V279', 'V280', 'V281', 'V282', 'V283', 'V284', 'V285'
'V289', 'V290', 'V291', 'V292', 'V293', 'V294', 'V295', 'V296', 'V297'
'V301', 'V302', 'V303', 'V304', 'V305', 'V306', 'V307', 'V308', 'V309'
'V313', 'V314', 'V315', 'V316', 'V317', 'V318', 'V319', 'V320', 'V321'
'V325', 'V326', 'V327', 'V328', 'V329', 'V330', 'V331', 'V332', 'V333'
'V337', 'V338', 'V339', 'id_01', 'id_02', 'id_03', 'id_04', 'id_05', 'id_06',
'id_09', 'id_10', 'id_11']
```

In [58]: `def analysis_num_fea(data, feature, target):`

```
    data_legit = data[data[target]==0][feature]
    data_fraud = data[data[target]==1][feature]

    sns.boxplot([data_legit, data_fraud], orient = 'h', palette = ['g','r'])
    plt.title(f'Box Plots of {feature} based on legit and fraud trasactions')
    plt.yticks([0,1],('Legit Transactions', 'Fraud Transactions'))
    plt.xlabel(f'{feature}')
    plt.grid()
    plt.show()

    legit = pd.Series(data_legit)
    fraud = pd.Series(data_fraud)

    x = PrettyTable()
    x.field_names = ["Feature Description", "Feature Description of legit transactions", "Feature Description of Fraud transactions"]
    x.add_row([data[feature].describe(), legit.describe(), fraud.describe()])
    print('\n')
    print('*Five Number Summary*')
    print(x)
    print('\n')
    print(f'The unique values in {feature} feature is:', len(data[feature].unique()))
    print(f'{feature} feature has {data[feature].isnull().sum()} null values, which is {data[feature].isnull().sum() / len(data[feature]) * 100} %')
    print('\n')

    y = PrettyTable()
    y.field_names = ["Percentile", "Legit Transactions", "Fraud Transactions"]

    for i in range(0,101,5):
        y.add_row([i, np.round(data_legit.quantile(i*0.01),3), np.round(data_fraud.quantile(i*0.01),3)])
    print('*Percentiles*')
    print(y)
    # First quartile (Q1)
    Q1_legit = np.round(data_legit.quantile(0.25),3)
    Q1_fraud = np.round(data_fraud.quantile(0.25),3)
    # Third quartile (Q3)
    Q3_legit = np.round(data_legit.quantile(0.75),3)
    Q3_fraud = np.round(data_fraud.quantile(0.75),3)
    print('InterQuartileRange of legit transactions:', Q3_legit - Q1_legit)
    print('InterQuartileRange of fraud transactions:', Q3_fraud - Q1_fraud)
    print('\n')

    plt.figure(figsize=(8,3))
    sns.distplot(data_legit, hist=True, label="Legit Transactions")
    sns.distplot(data_fraud, hist=True, label="Fraud Transactions")
    plt.title(f'{feature} of Legit Transaction and Fraud Transaction')
    plt.xlabel(f'{feature}')
    plt.legend()
```

```
plt.grid()
plt.show()
```

```
In [59]: def feature_analysis(df_train, feature, target):

    colour = ["#" + ''.join([random.choice('0123456789ABCDEF') for j in range(6)]) for i in range(10)]

    temp = pd.DataFrame(df_train.groupby(feature)[target].apply(np.mean)).reset_index()
    temp.columns = [feature, '% Fraud Transactions']
    plt.figure(figsize=(20,5))
    plt.bar(temp[feature], temp['% Fraud Transactions'], color = random.choice(colour))
    plt.xlabel(feature)
    plt.ylabel("% rate of fraud transactions")
    plt.xticks(rotation = 90)
    plt.show()
```

```
In [60]: def plot_features(data, feature, target):

    data_legit = data.loc[data[target]== 0]
    data_fraud = data.loc[data[target]== 1]

    x = PrettyTable()
    x.field_names = ["Feature Description", "Feature Description of legit transactions"]
    x.add_row([data[feature].describe(), data_legit[feature].describe(), data_fraud[feature].describe()])
    print('\n')
    print('*Five Number Summary*')
    print(x)
    print('\n')
    print(f'The unique values in {feature} feature is:', data[feature].nunique)
    print(f'{feature} feature has {data[feature].isnull().sum()} null values,')
    print('\n')
    feature_analysis(data, feature, target)
    univariate_barplots(data, feature, target)

    if (data[feature].dtype == 'int64') or (data[feature].dtype == float):
        plt.figure(figsize=(12,6))
        plt.subplot(1,2,1)
        sns.distplot(data_legit[feature], hist = False, label = 'legit transactions')
        sns.distplot(data_fraud[feature], hist = False, label = 'fraud transactions')
        plt.legend()
        plt.title(f'Analysis on {feature} feature')
        plt.grid()
        plt.subplot(1,2,2)
        sns.histplot(x=data[feature], hue=target, kde=True, data=data)
        plt.title(f'Analysis on {feature} feature')
        plt.grid()
        plt.show()
```

1.2.1 Feature: TransactionDT

```
In [61]: print(df_train['TransactionDT'])
# Sequential numbers beginning with 86400 make up the numbers in this feature.
# We are aware that this feature is a part of the date and time feature. and hence
# And hence we will encode this feature into datetime
```

```

0           86400
1           86401
2           86469
3           86499
4           86506
...
590535    15811047
590536    15811049
590537    15811079
590538    15811088
590539    15811131
Name: TransactionDT, Length: 590540, dtype: int64

```

```
In [62]: plt.figure(figsize=(6,4))
plt.hist(x = 'TransactionDT', data = df_train, label='Train')
plt.hist(x = 'TransactionDT', data = df_test, label='Test')
plt.title("Train vs Test TransactionDT")
plt.xlabel("TransactionDT")
plt.ylabel("Count")
plt.legend()
plt.show()
```



Observation: TransactionDT is all about seconds, as we know. We discovered that there is some lag in time between the train and test datasets after visualising the train and test distribution.

```
In [63]: # we will take a random start date and then append seconds to this date for our data
startdate = datetime.datetime.strptime('2021-12-31', '%Y-%m-%d')
df_train['TransactionDateTime'] = df_train['TransactionDT'].apply(lambda x: (startdate + timedelta(seconds=x)).strftime('%Y-%m-%d %H:%M:%S'))
df_test['TransactionDateTime'] = df_test['TransactionDT'].apply(lambda x: (startdate + timedelta(seconds=x)).strftime('%Y-%m-%d %H:%M:%S'))
```

```
In [64]: # feature engineering: converting datetime feature into repetitive day, month, year
df_train['TransactionDate'] = pd.DatetimeIndex(df_train['TransactionDateTime']).date
df_test['TransactionDate'] = pd.DatetimeIndex(df_test['TransactionDateTime']).date

df_train['TransactionDay'] = pd.DatetimeIndex(df_train['TransactionDateTime']).day
df_test['TransactionDay'] = pd.DatetimeIndex(df_test['TransactionDateTime']).day
```

```

df_train['TransactionMonth'] = pd.DatetimeIndex(df_train['TransactionDateTime']).month
df_test['TransactionMonth'] = pd.DatetimeIndex(df_test['TransactionDateTime']).month
df_train['TransactionMon_Yr'] = df_train['TransactionDateTime'].apply(lambda x: x.year * 100 + x.month)
df_train['TransactionYear'] = pd.DatetimeIndex(df_train['TransactionDateTime']).year
df_test['TransactionYear'] = pd.DatetimeIndex(df_test['TransactionDateTime']).year

df_train['TransactionWeekday'] = pd.DatetimeIndex(df_train['TransactionDateTime']).weekday
df_test['TransactionWeekday'] = pd.DatetimeIndex(df_test['TransactionDateTime']).weekday

df_train['TransactionHours'] = pd.DatetimeIndex(df_train['TransactionDateTime']).hour
df_test['TransactionHours'] = pd.DatetimeIndex(df_test['TransactionDateTime']).hour

```

1.2.1.1 TransactionDate

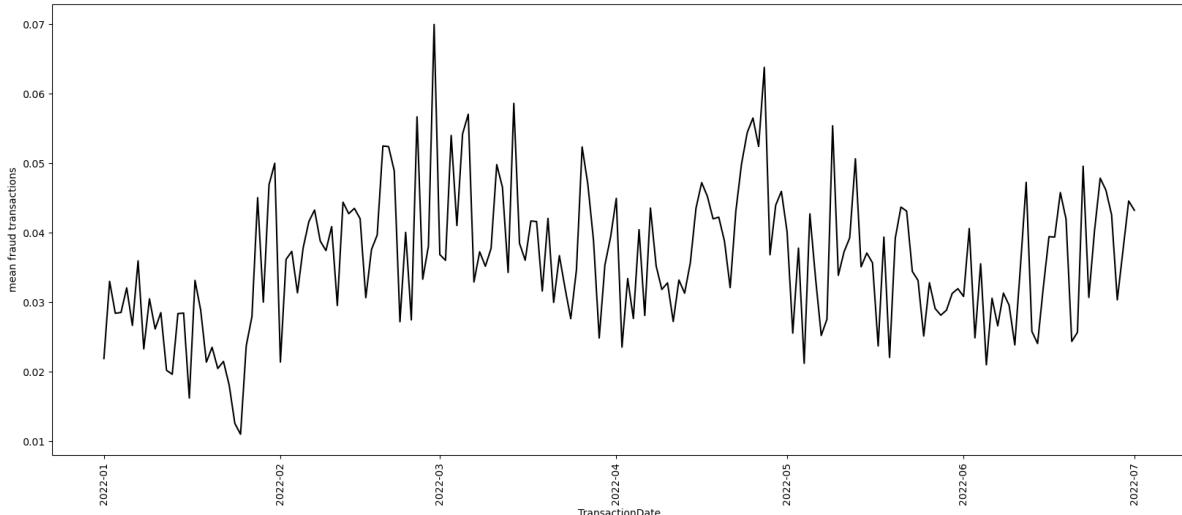
In [65]:

```

fig, ax = plt.subplots(figsize=(20, 8))

temp = pd.DataFrame(df_train.groupby('TransactionDate')['isFraud'].apply(np.mean))
temp.columns = ['TransactionDate', 'Mean Fraud']
ax.plot(temp['TransactionDate'], temp['Mean Fraud'], color = 'black')
plt.xlabel('TransactionDate')
plt.ylabel("mean fraud transactions")
plt.xticks(rotation = 90)
plt.show()

```

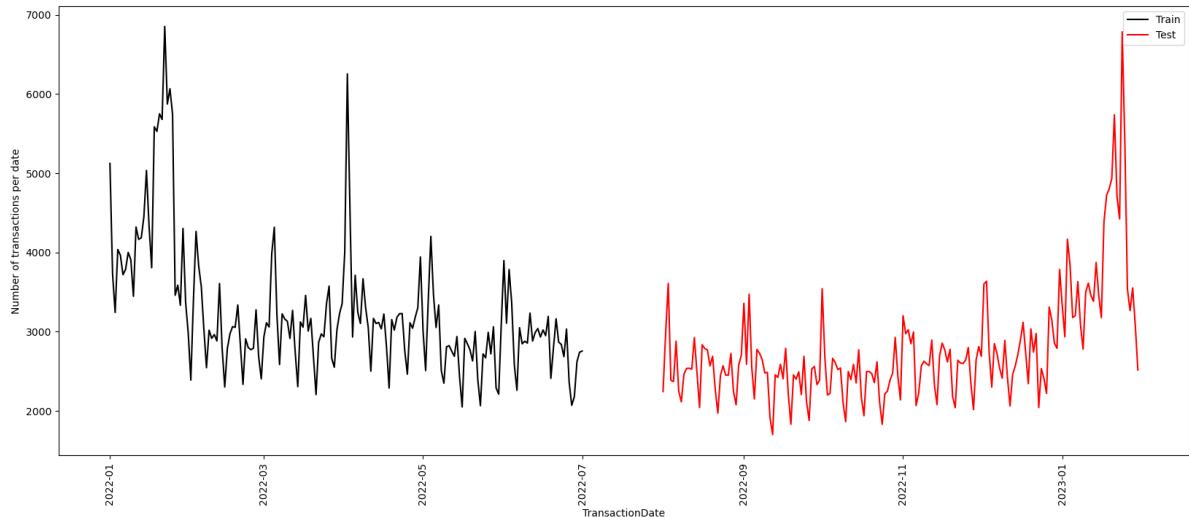


In [66]:

```

fig, axe = plt.subplots(1, 1, figsize=(20, 8))
axe.plot(df_train['TransactionDate'].value_counts().sort_index(), color = 'black')
axe.plot(df_test['TransactionDate'].value_counts().sort_index(), color= 'red')
plt.legend(['Train', 'Test'])
plt.xlabel('TransactionDate')
plt.ylabel("Number of transactions per date")
plt.xticks(rotation = 90)
plt.show()

```



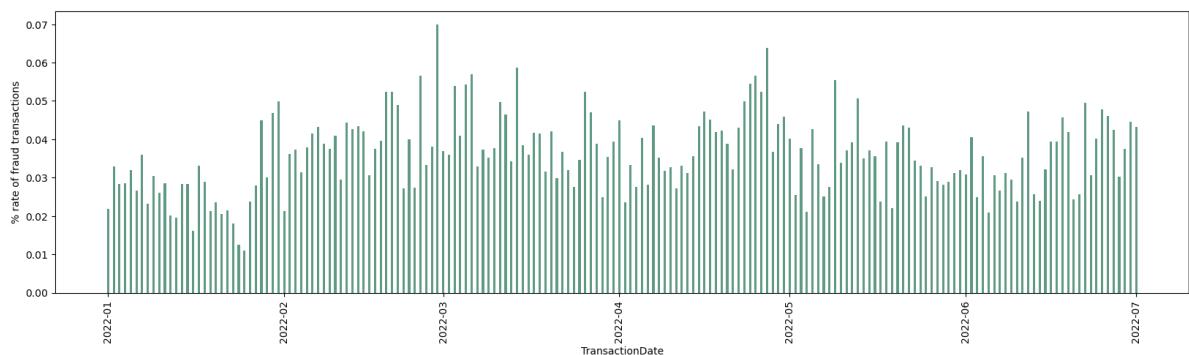
Observation: From the above plot we confirm that there is a gap of aprrox 30 days between our train and test dataset.

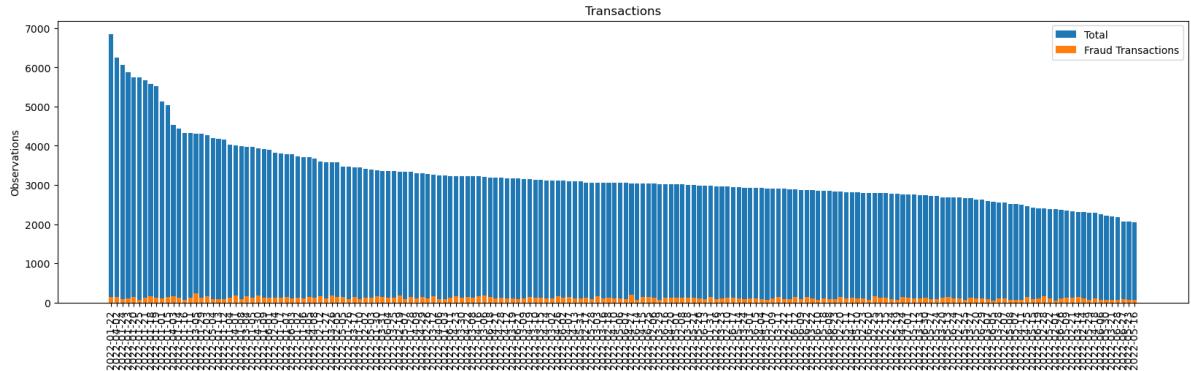
```
In [67]: plot_features(df_train, 'TransactionDate', 'isFraud')
```

Five Number Summary

Feature Description		Feature Description of legit transactions	
Feature Description of fraud transactions			
count	590540	count	569877
count	20663		
unique	182	unique	182
unique	182		
top	2022-01-22	top	2022-01-22
top	2022-03-05		
freq	6852	freq	6705
freq	234		
Name: TransactionDate, dtype: object		Name: TransactionDate, dtype: object	
Name: TransactionDate, dtype: object			

The unique values in TransactionDate feature is: 182
 TransactionDate feature has 0 null values, which holds 0.0 %





	TransactionDate	isFraud	total	% Occurence	Avg Fraud
21	2022-01-22	147	6852	0.011603	0.021454
91	2022-04-02	147	6252	0.010587	0.023512
23	2022-01-24	76	6065	0.010270	0.012531
22	2022-01-23	106	5872	0.009943	0.018052
19	2022-01-20	135	5749	0.009735	0.023482
..
79	2022-03-21	66	2204	0.003732	0.029946
178	2022-06-28	66	2177	0.003686	0.030317
177	2022-06-27	88	2068	0.003502	0.042553
142	2022-05-23	71	2064	0.003495	0.034399
135	2022-05-16	73	2048	0.003468	0.035645

[182 rows x 5 columns]

Observation: There are some days where there were more transactions, but no clear inference can be drawn from this.

1.2.1.2 Other Time features ('TransactionDay', 'TransactionMon_Yr', 'TransactionWeekday', 'TransactionHours')

```
In [68]: timefea = ['TransactionDay', 'TransactionMon_Yr', 'TransactionWeekday', 'Transacti
```

```
In [69]: for fea in timefea:
    print(f'Analysis of {fea} feature')
    plot_features(df_train, fea, 'isFraud')
    print('\n')
    print('*'*120)
    print('\n')
```

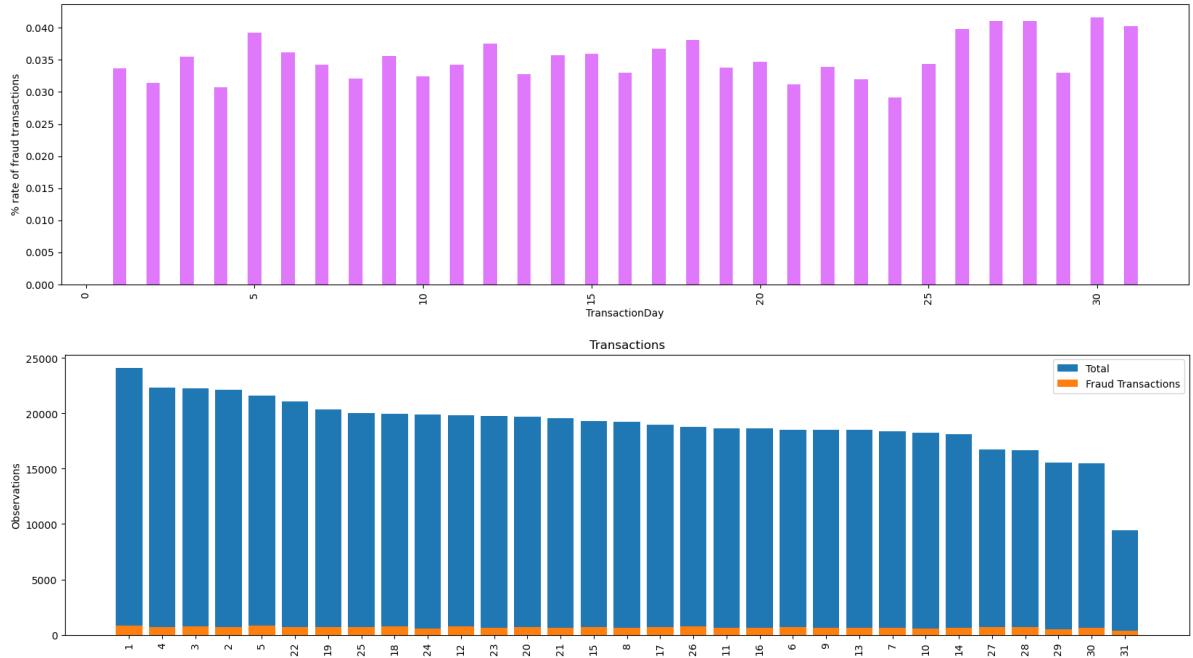
Analysis of TransactionDay feature

Five Number Summary

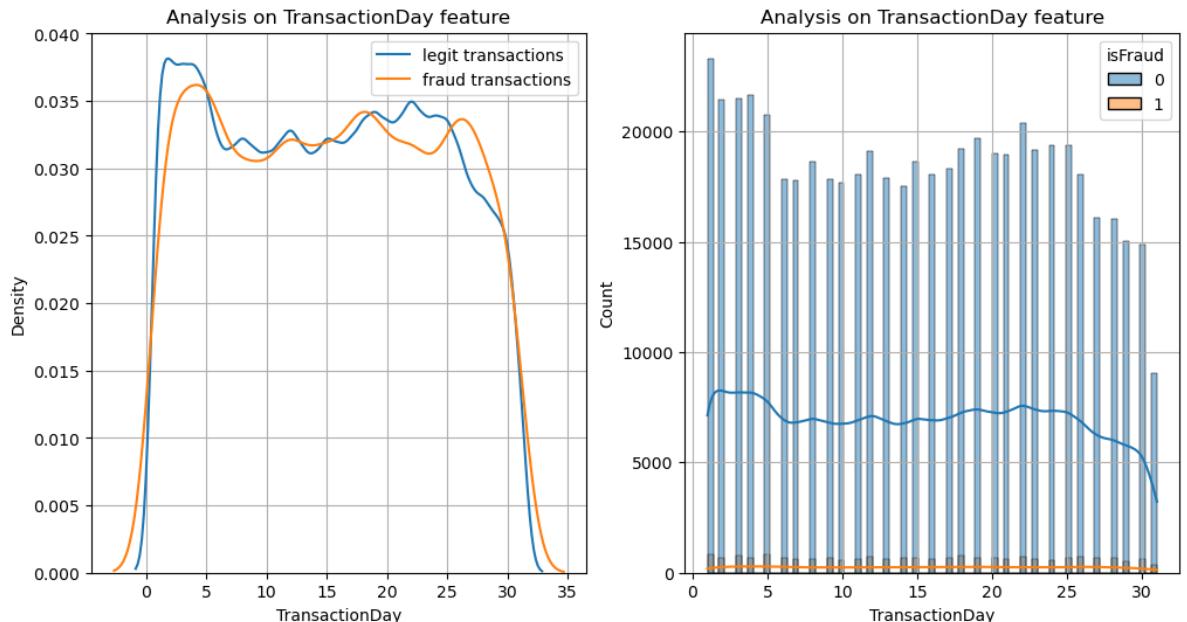
Feature Description		Feature Description of legit transactions	
Feature Description of fraud transactions			
count	590540.000000	count	569877.000000
count	20663.000000	mean	15.212521
mean	15.221612	mean	15.472342
std	8.823652	std	8.820742
std	8.900100	min	1.000000
min	1.000000	min	1.000000
25%	7.000000	25%	7.000000
25%	8.000000	50%	15.000000
50%	15.000000	50%	15.000000
75%	23.000000	75%	23.000000
75%	23.000000	max	31.000000
max	31.000000	max	31.000000
max	31.000000		
Name: TransactionDay, dtype: float64		Name: TransactionDay, dtype: float64	
Name: TransactionDay, dtype: float64			

The unique values in TransactionDay feature is: 31

TransactionDay feature has 0 null values, which holds 0.0 %



	TransactionDay	isFraud	total	% Occurrence	Avg Fraud
0		1	811	0.040836	0.033631
3		4	687	0.037806	0.030771
2		3	789	0.037693	0.035446
1		2	695	0.037437	0.031437
4		5	848	0.036585	0.039250
21		22	715	0.035738	0.033878
18		19	688	0.034472	0.033797
24		25	688	0.033938	0.034328
17		18	761	0.033813	0.038111
23		24	581	0.033747	0.029153
11		12	745	0.033588	0.037560
22		23	632	0.033474	0.031971
19		20	684	0.033368	0.034712
20		21	610	0.033112	0.031196
14		15	696	0.032743	0.035995
7		8	617	0.032609	0.032040
16		17	697	0.032164	0.036696
25		26	748	0.031786	0.039849
10		11	640	0.031612	0.034283
15		16	615	0.031563	0.032995
5		6	670	0.031343	0.036199
8		9	659	0.031337	0.035610
12		13	605	0.031309	0.032722
6		7	631	0.031156	0.034295
9		10	592	0.030904	0.032438
13		14	648	0.030718	0.035722
26		27	688	0.028372	0.041062
27		28	686	0.028278	0.041080
28		29	513	0.026298	0.033033
29		30	644	0.026223	0.041586
30		31	380	0.015979	0.040271



```
*****
*****
```

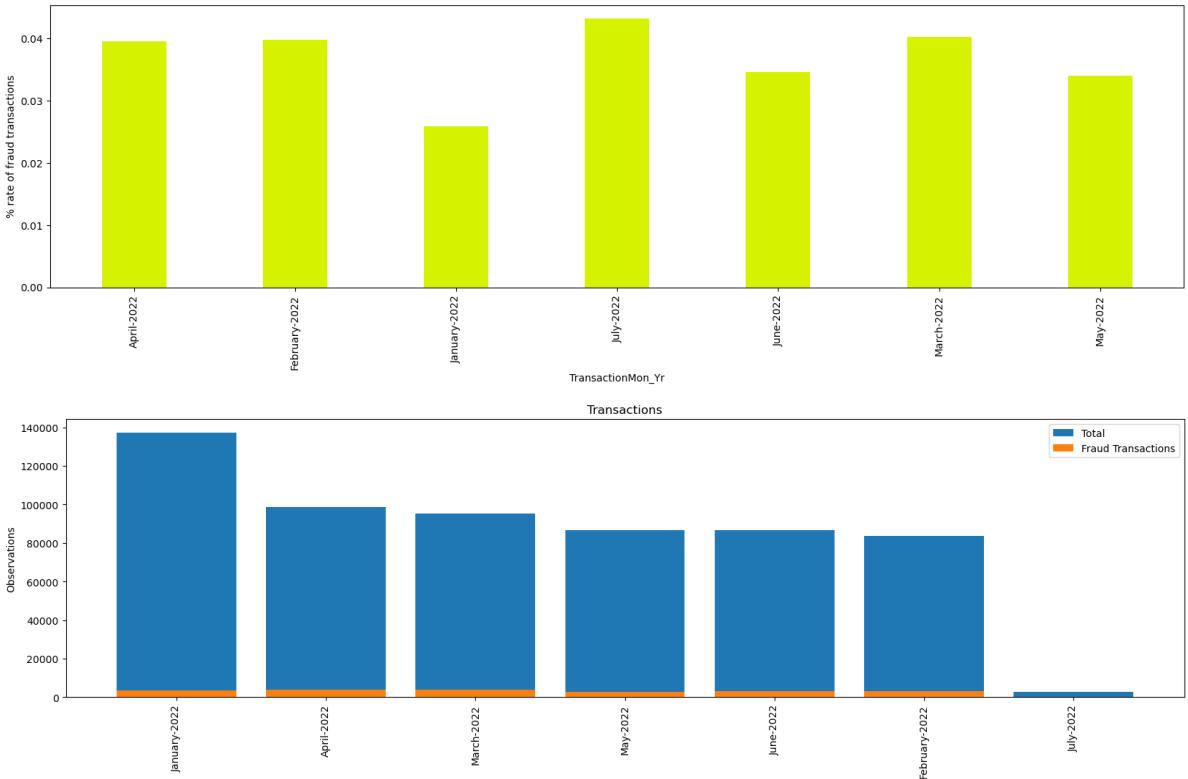
Analysis of TransactionMon_Yr feature

Five Number Summary

Feature Description		Feature Description of legit transactions	
Feature Description of fraud transactions			
count	590540	count	569877
count	20663	unique	7
unique	7	unique	7
unique	7	top	January-2022
top	January-2022	top	January-2022
top	April-2022	freq	133771
freq	137321	freq	133771
freq	3898		
Name: TransactionMon_Yr, dtype: object		Name: TransactionMon_Yr, dtype: object	
Name: TransactionMon_Yr, dtype: object			

The unique values in TransactionMon_Yr feature is: 7

TransactionMon_Yr feature has 0 null values, which holds 0.0 %



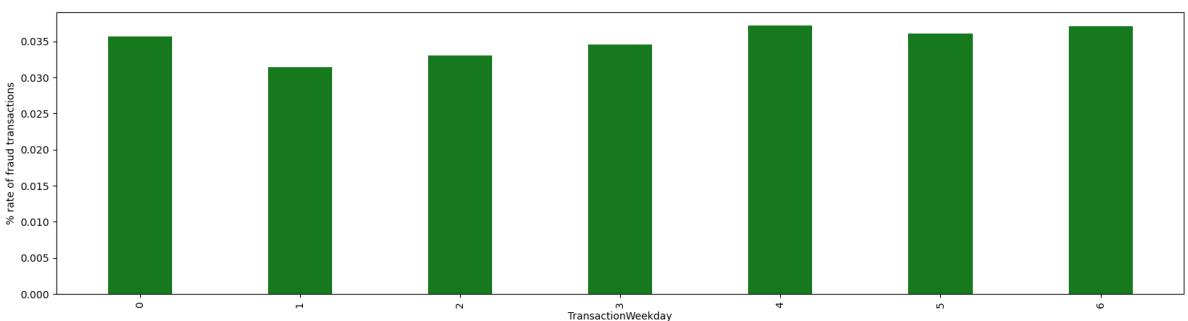
	TransactionMon_Yr	isFraud	total	% Occurence	Avg Fraud
2	January-2022	3550	137321	0.232535	0.025852
0	April-2022	3898	98615	0.166991	0.039527
5	March-2022	3832	95121	0.161075	0.040286
6	May-2022	2949	86672	0.146767	0.034025
4	June-2022	2995	86572	0.146598	0.034595
1	February-2022	3320	83485	0.141371	0.039768
3	July-2022	119	2754	0.004664	0.043210

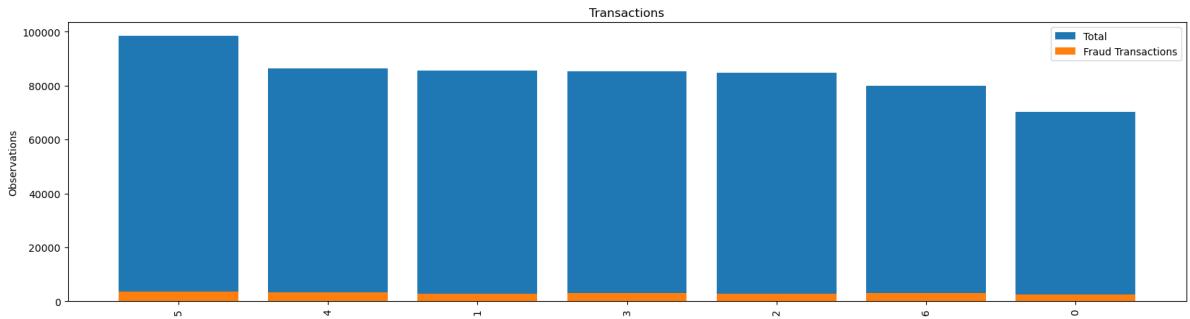
Analysis of TransactionWeekday feature

Five Number Summary

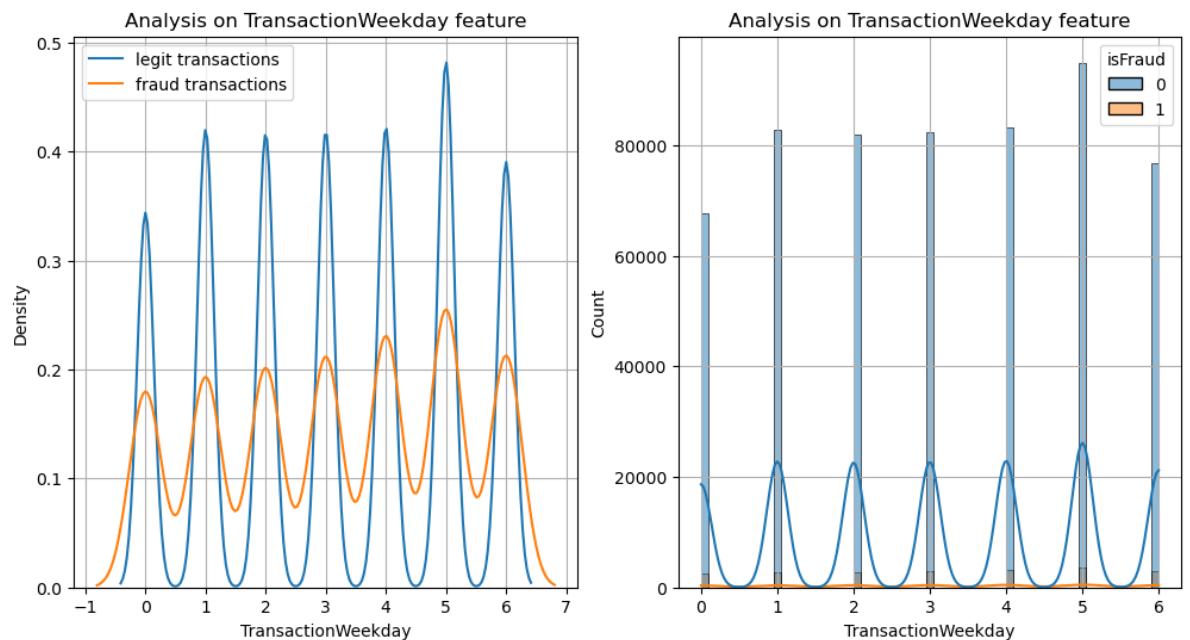
Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	590540.000000	count	569877.000000
count	20663.000000	mean	3.093036
mean	3.095731	std	1.952429
mean	3.170062	min	0.000000
std	1.952824	25%	1.000000
std	1.962261	25%	1.000000
min	0.000000	50%	3.000000
min	0.000000	50%	3.000000
25%	1.000000	75%	5.000000
25%	1.000000	75%	5.000000
50%	3.000000	max	6.000000
50%	3.000000	max	6.000000
75%	5.000000		
75%	5.000000		
max	6.000000		
max	6.000000		
Name: TransactionWeekday, dtype: float64		Name: TransactionWeekday, dtype: float64	
Name: TransactionWeekday, dtype: float64		Name: TransactionWeekday, dtype: float64	

The unique values in TransactionWeekday feature is: 7
 TransactionWeekday feature has 0 null values, which holds 0.0 %





TransactionWeekday	isFraud	total	% Occurrence	Avg Fraud
5	5	3550	0.166800	0.036040
4	4	3211	0.146268	0.037174
1	1	2687	0.144669	0.031452
3	3	2946	0.144539	0.034514
2	2	2803	0.143623	0.033048
6	6	2963	0.135188	0.037115
0	0	2503	0.118913	0.035644



```
*****
*****
```

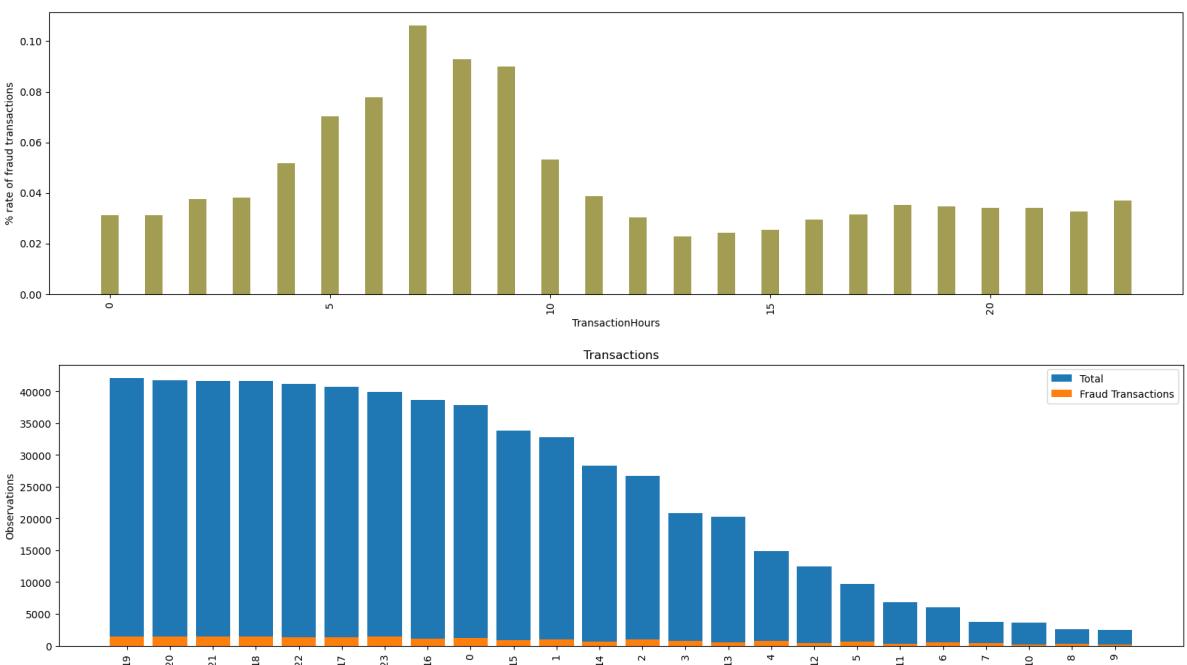
Analysis of TransactionHours feature

Five Number Summary

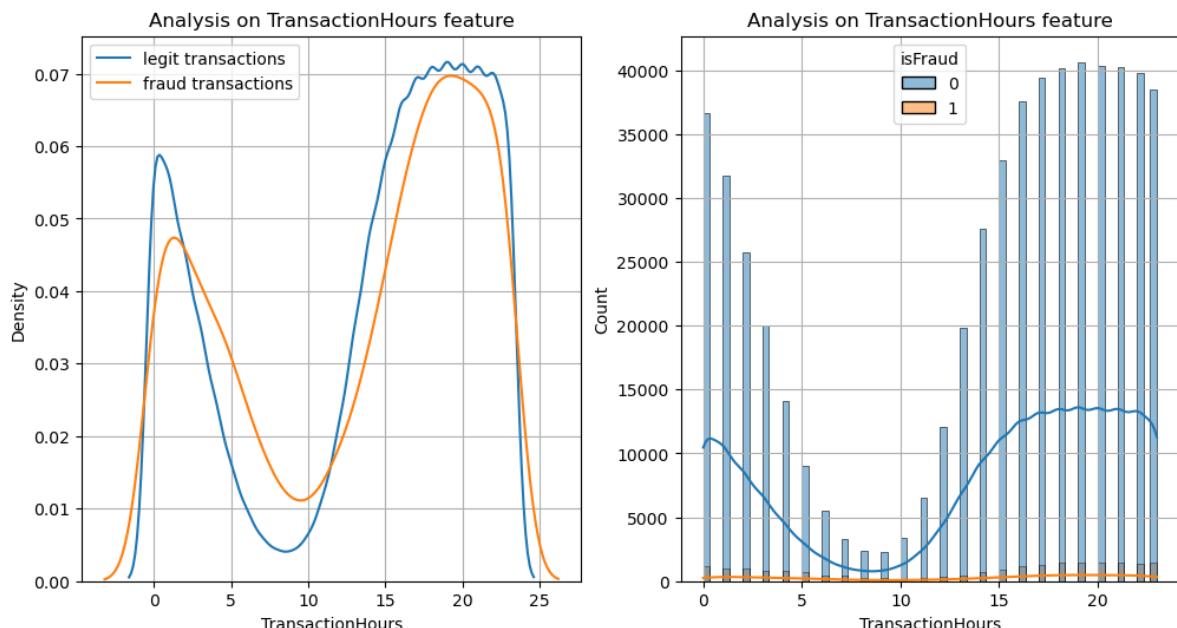
	Feature Description	Feature Description of legit transactions	
	Feature Description of fraud transactions		
count	590540.000000	count	569877.000000
count	20663.000000		
mean	13.861923	mean	13.880916
mean	13.338092		
std	7.607152	std	7.600842
std	7.761035		
min	0.000000	min	0.000000
min	0.000000		
25%	6.000000	25%	6.000000
25%	5.000000		
50%	16.000000	50%	16.000000
50%	16.000000		
75%	20.000000	75%	20.000000
75%	20.000000		
max	23.000000	max	23.000000
max	23.000000		
Name: TransactionHours, dtype: float64		Name: TransactionHours, dtype: float64	
4 Name: TransactionHours, dtype: float64			

The unique values in TransactionHours feature is: 24

TransactionHours feature has 0 null values, which holds 0.0 %



	TransactionHours	isFraud	total	% Occurence	Avg Fraud
19	19	1463	42115	0.071316	0.034738
20	20	1432	41782	0.070752	0.034273
21	21	1416	41641	0.070513	0.034005
18	18	1467	41639	0.070510	0.035231
22	22	1345	41139	0.069663	0.032694
17	17	1284	40723	0.068959	0.031530
23	23	1478	39949	0.067648	0.036997
16	16	1142	38698	0.065530	0.029511
0	0	1186	37795	0.064001	0.031380
15	15	860	33859	0.057336	0.025399
1	1	1027	32797	0.055537	0.031314
14	14	686	28328	0.047970	0.024216
2	2	1002	26732	0.045267	0.037483
3	3	797	20802	0.035225	0.038314
13	13	465	20315	0.034401	0.022889
4	4	770	14839	0.025128	0.051890
12	12	379	12451	0.021084	0.030439
5	5	682	9701	0.016427	0.070302
11	11	265	6827	0.011561	0.038816
6	6	467	6007	0.010172	0.077743
7	7	393	3704	0.006272	0.106102
10	10	193	3627	0.006142	0.053212
8	8	241	2591	0.004388	0.093014
9	9	223	2479	0.004198	0.089956

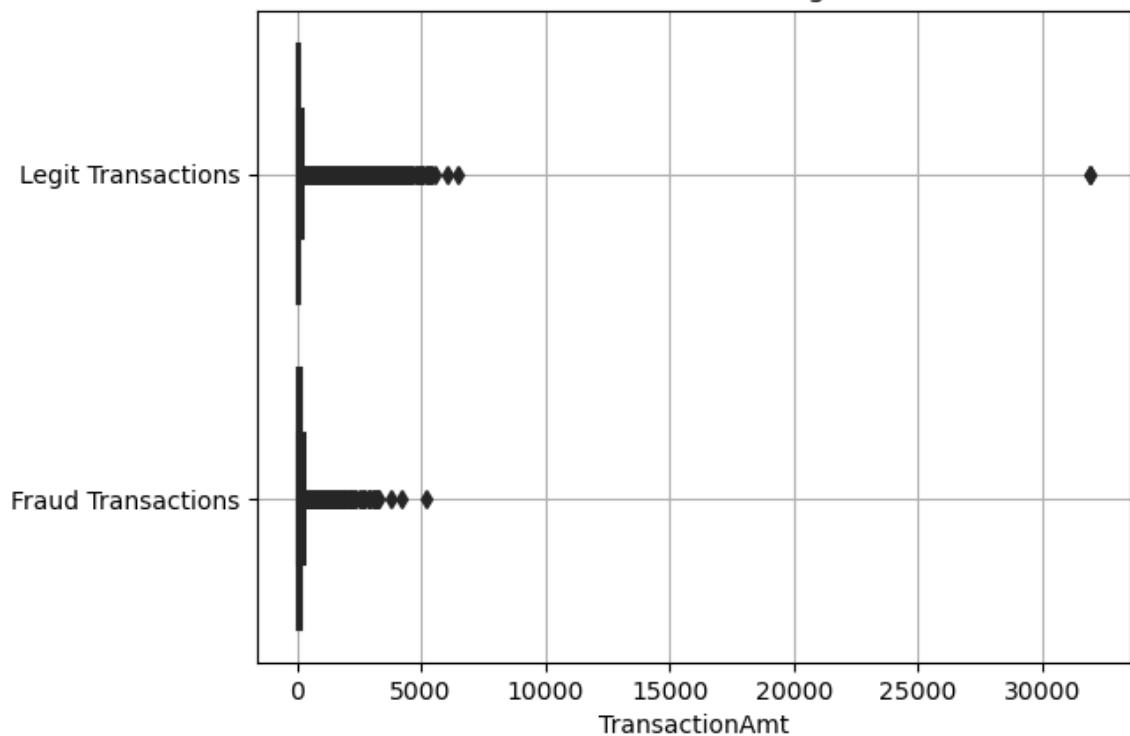


```
*****
*****
```

Observation: There are some days, months and hours where more transactions or fraud transactions have been seen, but this information cannot be used to make any firm conclusions.

1.2.2 Feature: TransactionAmt

```
In [70]: analysis_num_fea(df_train, 'TransactionAmt', 'isFraud')
```

Box Plots of TransactionAmt based on legit and fraud trasactions

Five Number Summary

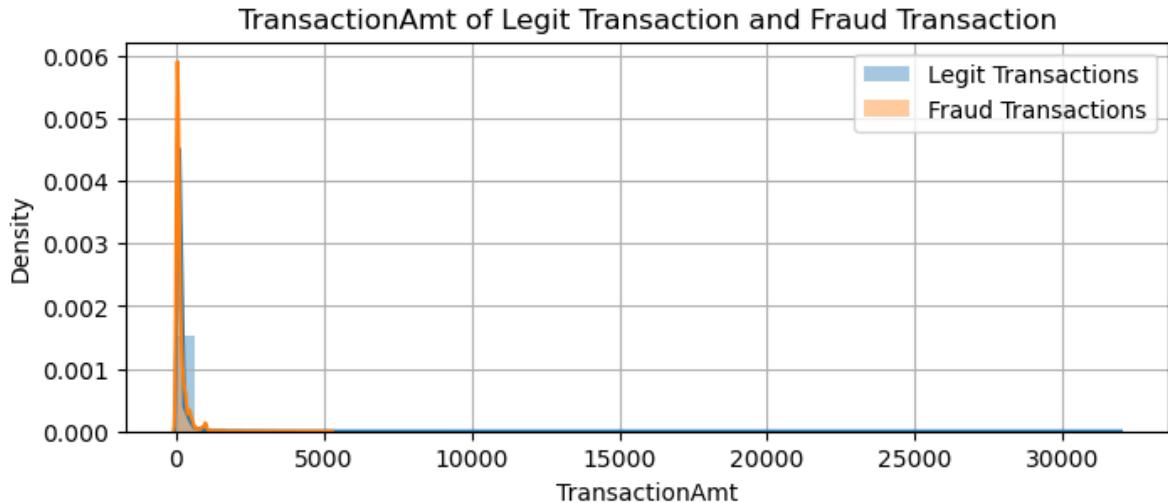
Feature Description		Feature Description of legit transactions	
Feature Description of fraud transactions			
count	590540.000000	count	569877.000000
count	20663.000000		
mean	135.027176	mean	134.511665
mean	149.244779		
std	239.162522	std	239.395078
std	232.212163		
min	0.251000	min	0.251000
min	0.292000		
25%	43.321000	25%	43.970000
25%	35.044000		
50%	68.769000	50%	68.500000
50%	75.000000		
75%	125.000000	75%	120.000000
75%	161.000000		
max	31937.391000	max	31937.391000
max	5191.000000		
Name: TransactionAmt, dtype: float64		Name: TransactionAmt, dtype: float64	
Name: TransactionAmt, dtype: float64			

The unique values in TransactionAmt feature is: 20902
 TransactionAmt feature has 0 null values, which holds 0.0 %

Percentiles

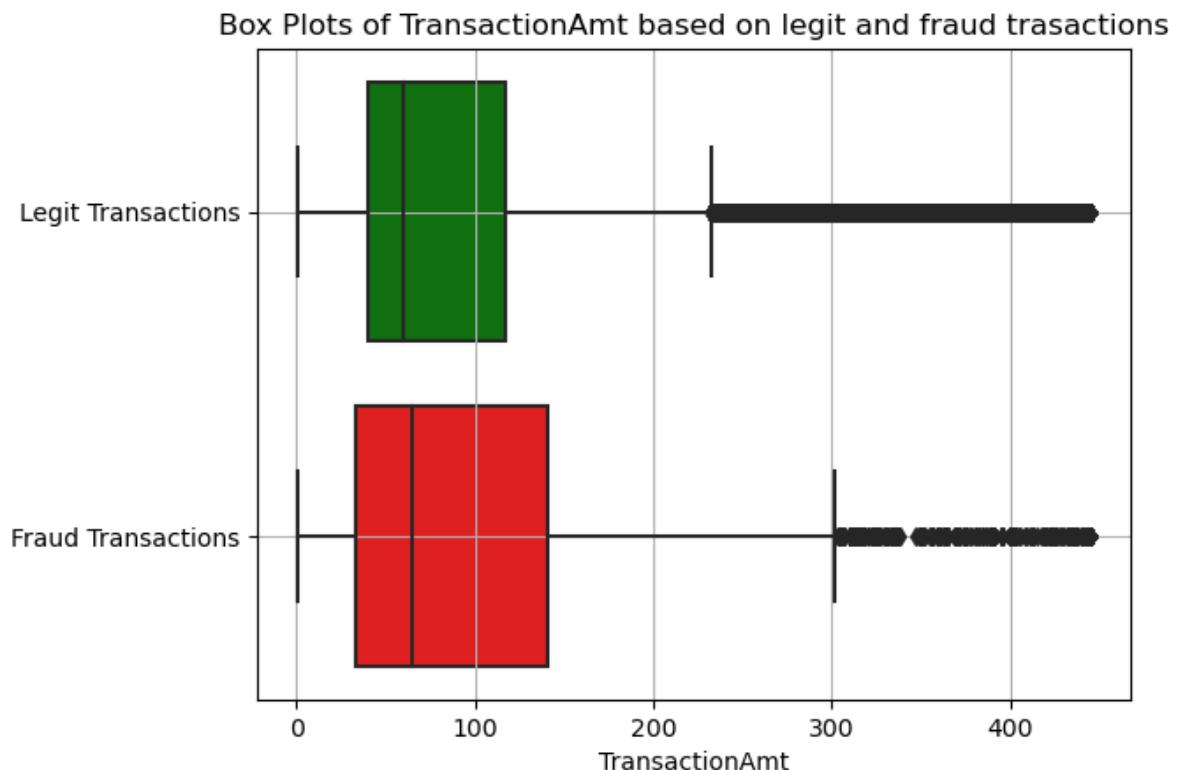
Percentile	Legit Transactions	Fraud Transactions
0	0.251	0.292
5	20.73	13.063
10	26.31	18.936
15	30.95	25.0
20	35.95	29.818
25	43.97	35.044
30	49.0	42.222
35	52.95	49.0
40	57.95	55.0
45	59.0	59.0
50	68.5	75.0
55	83.5	87.0
60	100.0	100.0
65	107.95	117.0
70	117.0	141.0
75	120.0	161.0
80	159.95	201.0
85	209.95	262.263
90	267.112	335.0
95	435.0	500.0
100	31937.391	5191.0

InterQuartileRange of legit transactions: 76.03
 InterQuartileRange of fraud transactions: 125.956



Observation: The TransactionAmt feature is significantly impacted by outliers and is highly right-skewed, as shown by the box plot. Since the percentiles show a significant difference in TransactionAmt between the 95th and 100th percentiles, we will use values till 95th percentiles for our analysis.

```
In [71]: analysis_num_fea(df_train[df_train['TransactionAmt'] <= np.percentile(df_train['Transac
```



Five Number Summary

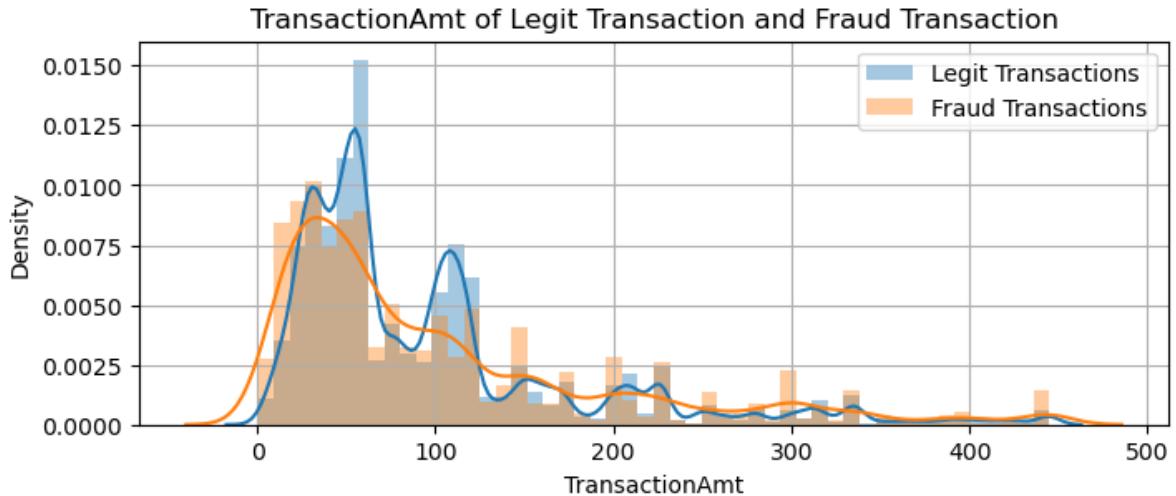
Feature Description		Feature Description of legit transactions	
Feature Description of fraud transactions			
count	563535.000000	count	544231.000000
count	19304.000000		
mean	96.909680	mean	96.713123
mean	102.451155		
std	83.502508	std	82.977458
std	96.982891		
min	0.251000	min	0.251000
min	0.292000		
25%	40.000000	25%	40.110000
25%	33.593000		
50%	59.950000	50%	59.630000
50%	64.713000		
75%	117.000000	75%	117.000000
75%	141.000000		
max	445.000000	max	445.000000
max	445.000000		
Name: TransactionAmt, dtype: float64		Name: TransactionAmt, dtype: float64	
Name: TransactionAmt, dtype: float64			

The unique values in TransactionAmt feature is: 13809
 TransactionAmt feature has 0 null values, which holds 0.0 %

Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.251	0.292
5	20.0	12.709
10	25.95	18.198
15	30.95	24.346
20	35.95	29.0
25	40.11	33.593
30	48.97	39.0
35	50.0	46.757
40	57.95	50.0
45	59.0	59.0
50	59.63	64.713
55	77.0	77.0
60	92.0	90.475
65	100.0	100.0
70	107.95	117.0
75	117.0	141.0
80	135.5	160.331
85	171.0	200.0
90	211.0	250.0
95	281.8	311.95
100	445.0	445.0

InterQuartileRange of legit transactions: 76.89
 InterQuartileRange of fraud transactions: 107.407

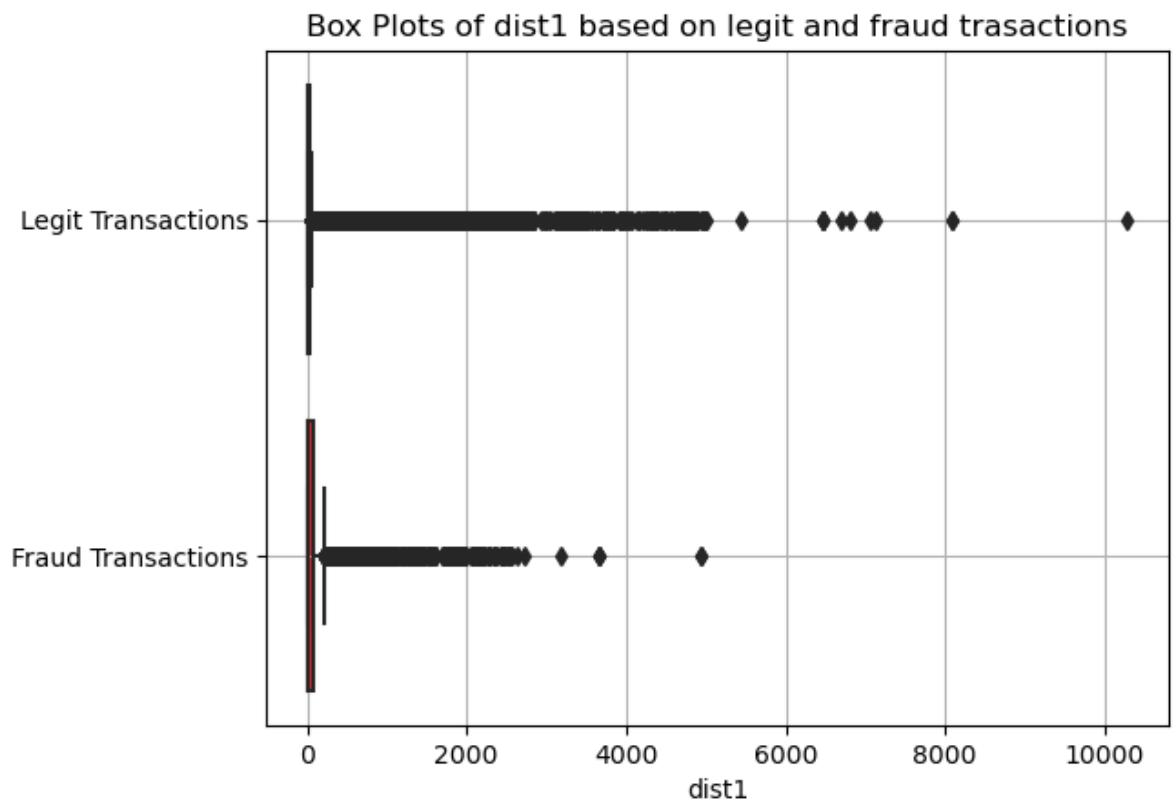


Observation: Distribution of legitimate transactions versus fraudulent transactions varies greatly. Most of the differences are between the 65th and 95th percentiles. A fraudulent transaction appears to have a greater amount than a legitimate transaction.

1.2.3 Feature: 'distance features'

```
In [72]: dist_fea = ['dist1', 'dist2']
for fea in dist_fea:
    print(f'Analysis of {fea} feature')
    analysis_num_fea(df_train, fea, 'isFraud')
    print('\n')
    print('*'*120)
    print('\n')
```

Analysis of dist1 feature



Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	238269.000000	count	233514.000000	
count	4755.000000			
mean	118.502180	mean	117.360098	
mean	174.588854			
std	371.872026	std	370.641879	
std	424.232313			
min	0.000000	min	0.000000	
min	0.000000			
25%	3.000000	25%	3.000000	
25%	3.000000			
50%	8.000000	50%	8.000000	
50%	10.000000			
75%	24.000000	75%	24.000000	
75%	85.500000			
max	10286.000000	max	10286.000000	
max	4942.000000			
Name: dist1, dtype: float64		Name: dist1, dtype: float64		
Name: dist1, dtype: float64				

The unique values in dist1 feature is: 2652

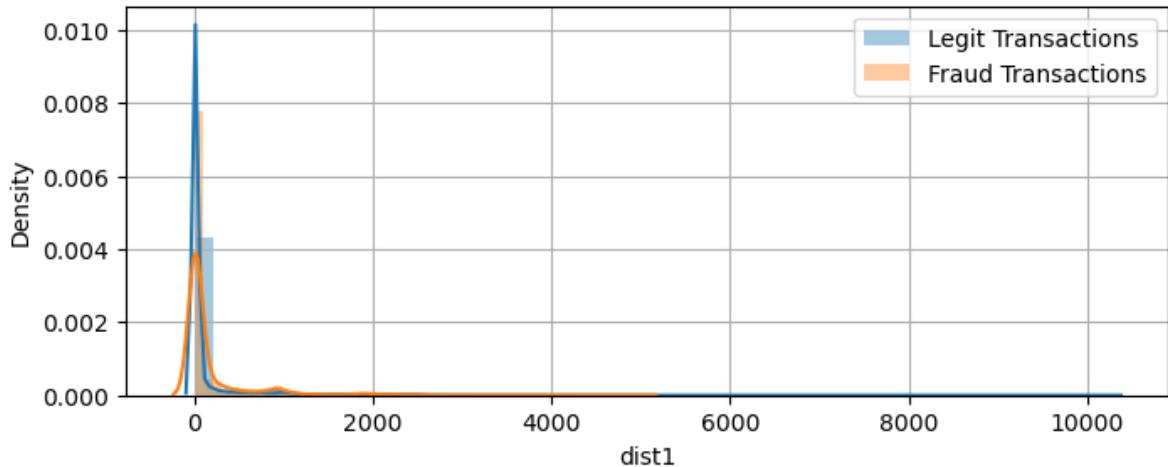
dist1 feature has 352271 null values, which holds 0.596523520845328 %

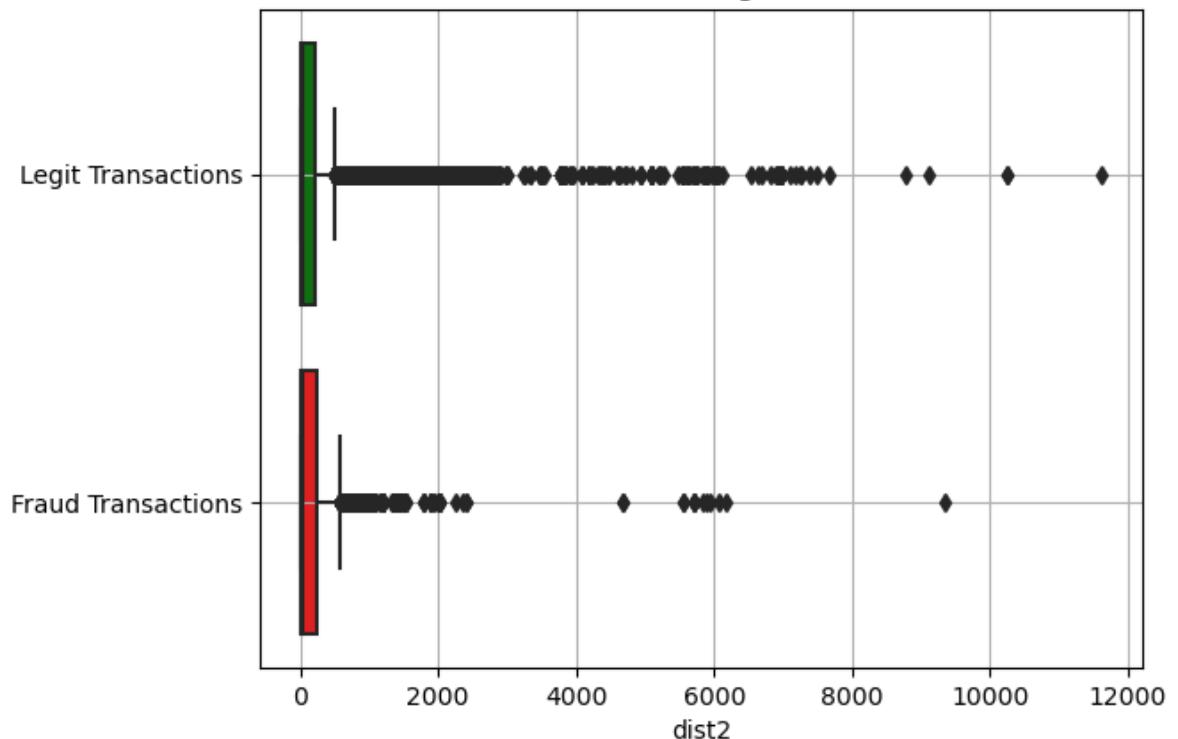
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	1.0	1.0
15	1.0	1.0
20	2.0	2.0
25	3.0	3.0
30	4.0	4.0
35	5.0	5.0
40	6.0	7.0
45	7.0	9.0
50	8.0	10.0
55	10.0	14.0
60	12.0	18.0
65	14.0	23.0
70	18.0	35.8
75	24.0	85.5
80	35.0	167.0
85	75.0	289.0
90	255.0	642.0
95	837.0	976.0
100	10286.0	4942.0

InterQuartileRange of legit transactions: 21.0

InterQuartileRange of fraud transactions: 82.5

dist1 of Legit Transaction and Fraud Transaction

Analysis of dist2 feature**Box Plots of dist2 based on legit and fraud trasactions**

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	37627.000000	count	33896.000000	count
mean	231.855423	mean	235.199758	mean
std	529.053494	std	536.798492	std
min	0.000000	min	0.000000	min
25%	7.000000	25%	7.000000	25%
50%	37.000000	50%	36.000000	50%
75%	49.000000	75%	206.000000	75%
max	233.000000	max	11623.000000	max
	11623.000000		11623.000000	
	Name: dist2, dtype: float64		Name: dist2, dtype: float64	
	Name: dist2, dtype: float64			

The unique values in dist2 feature is: 1752

dist2 feature has 552913 null values, which holds 0.9362837403054831 %

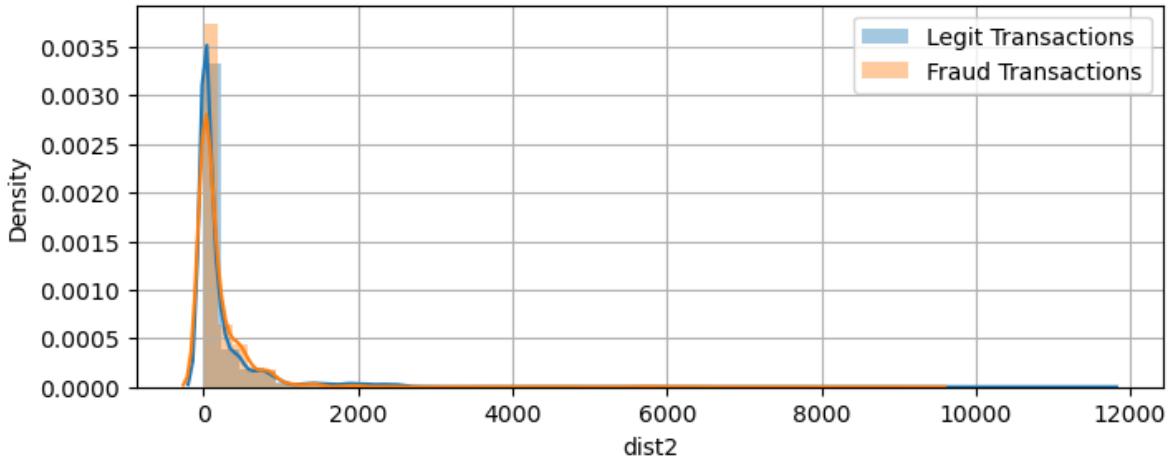
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	1.0	0.0
15	3.0	1.0
20	7.0	2.0
25	7.0	7.0
30	7.0	7.0
35	9.0	7.0
40	14.0	10.0
45	23.0	18.0
50	36.0	49.0
55	62.0	68.0
60	84.0	125.0
65	128.0	144.0
70	158.0	188.0
75	206.0	233.0
80	301.0	321.0
85	438.0	442.0
90	651.0	552.0
95	1069.0	810.0
100	11623.0	9337.0

InterQuartileRange of legit transactions: 199.0

InterQuartileRange of fraud transactions: 226.0

dist2 of Legit Transaction and Fraud Transaction



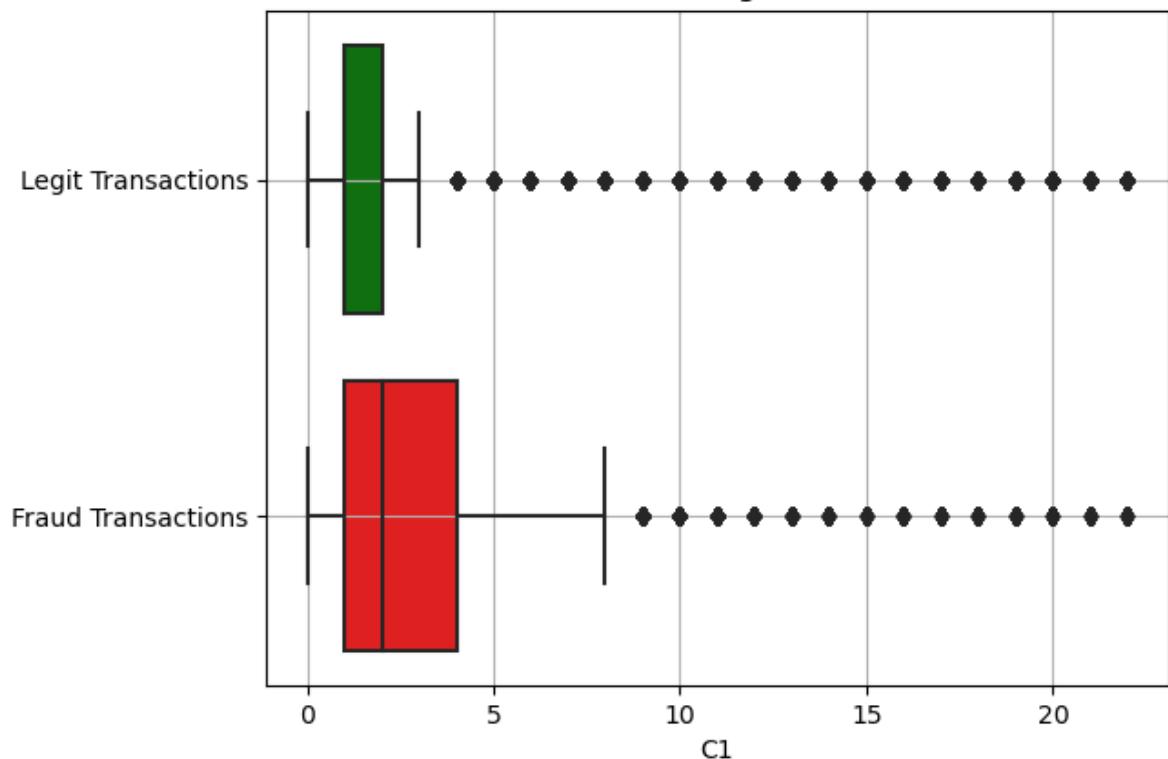
Observation: The average distance('dist1' & 'dist2') seems to be on a higher side for fraud transactions.

1.2.4 Feature: 'C features'

```
In [73]: c_features = ['C1', 'C2', 'C3', 'C4', 'C5', 'C6', 'C7', 'C8', 'C9', 'C10', 'C11',  
# These features are impacted by outliers and are right-skewed,  
# we will use values till 95th percentiles for our analysis.  
for fea in c_features:  
    print(f'Analysis of {fea} feature')  
    analysis_num_fea(df_train[df_train[fea]<=np.percentile(df_train[fea], 95)], fea)  
    print('\n')  
    print('*'*120)  
    print('\n')
```

Analysis of C1 feature

Box Plots of C1 based on legit and fraud transactions



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	561119.000000	count	542848.000000	18271.000000
mean	2.255258	mean	2.212347	3.530184
std	2.454223	std	2.367579	4.093483
min	0.000000	min	0.000000	0.000000
25%	1.000000	25%	1.000000	1.000000
50%	2.000000	50%	1.000000	2.000000
75%	2.000000	75%	2.000000	4.000000
max	22.000000	max	22.000000	22.000000
Name: C1, dtype: float64		Name: C1, dtype: float64		Name: C1, dtype: float64

The unique values in C1 feature is: 23

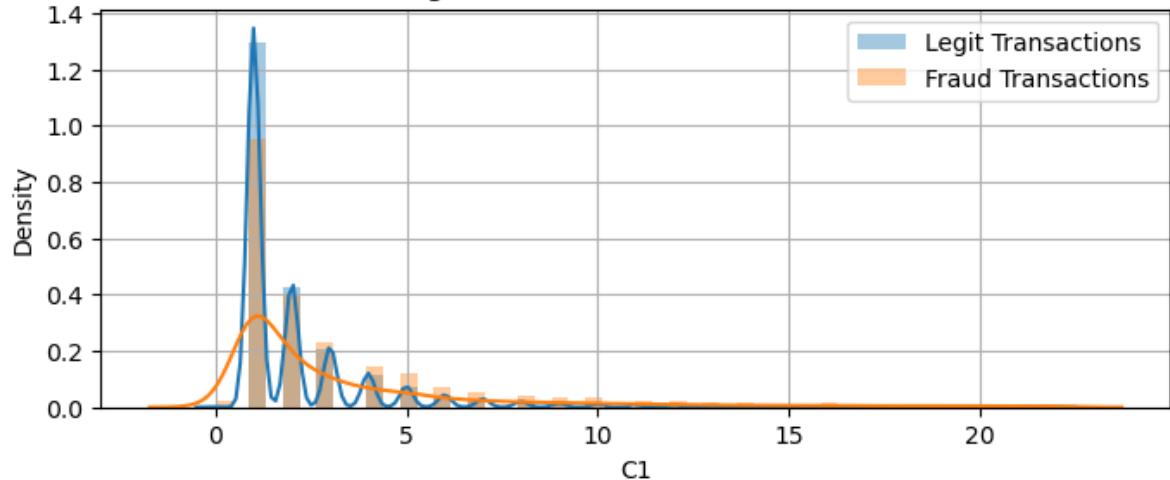
C1 feature has 0 null values, which holds 0.0 %

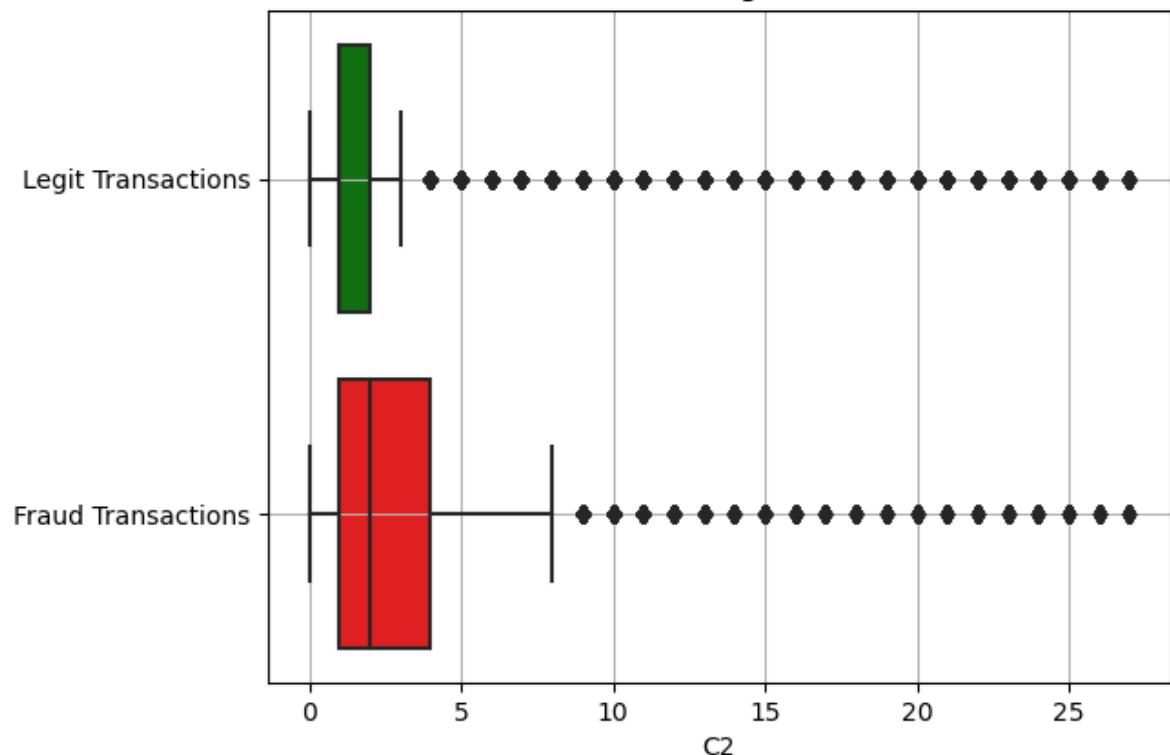
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	1.0	1.0
10	1.0	1.0
15	1.0	1.0
20	1.0	1.0
25	1.0	1.0
30	1.0	1.0
35	1.0	1.0
40	1.0	1.0
45	1.0	2.0
50	1.0	2.0
55	1.0	2.0
60	2.0	2.0
65	2.0	3.0
70	2.0	3.0
75	2.0	4.0
80	3.0	5.0
85	4.0	6.0
90	5.0	9.0
95	7.0	13.0
100	22.0	22.0

InterQuartileRange of legit transactions: 1.0

InterQuartileRange of fraud transactions: 3.0

C1 of Legit Transaction and Fraud Transaction

Analysis of C2 feature**Box Plots of C2 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	561093.000000	count	543013.000000	
count	18080.000000			
mean	2.330587	mean	2.281264	
mean	3.811947			
std	2.737550	std	2.633049	
std	4.699362			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	2.000000			
75%	3.000000	75%	2.000000	
75%	4.000000			
max	27.000000	max	27.000000	
max	27.000000			
Name: C2, dtype: float64		Name: C2, dtype: float64		
Name: C2, dtype: float64				

The unique values in C2 feature is: 28

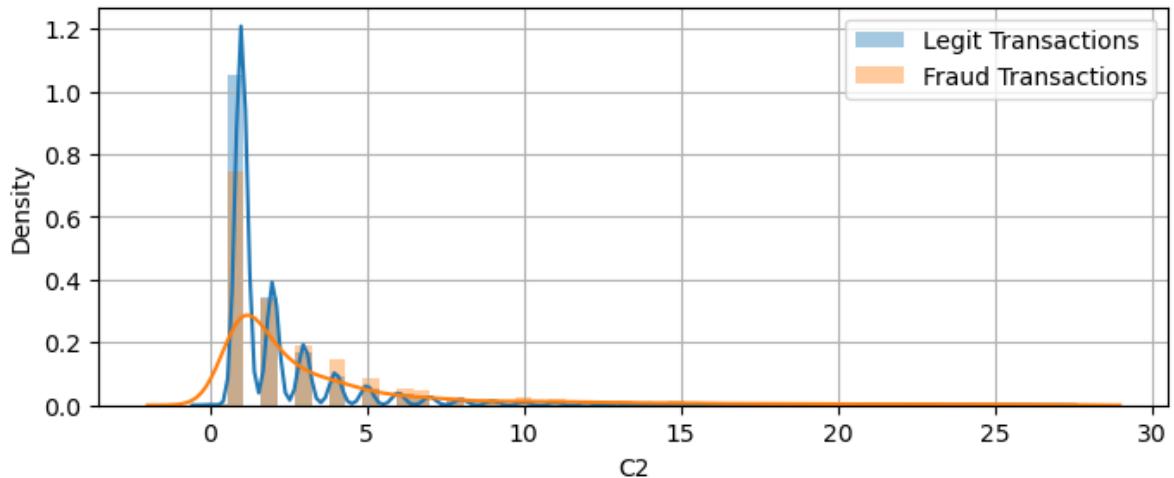
C2 feature has 0 null values, which holds 0.0 %

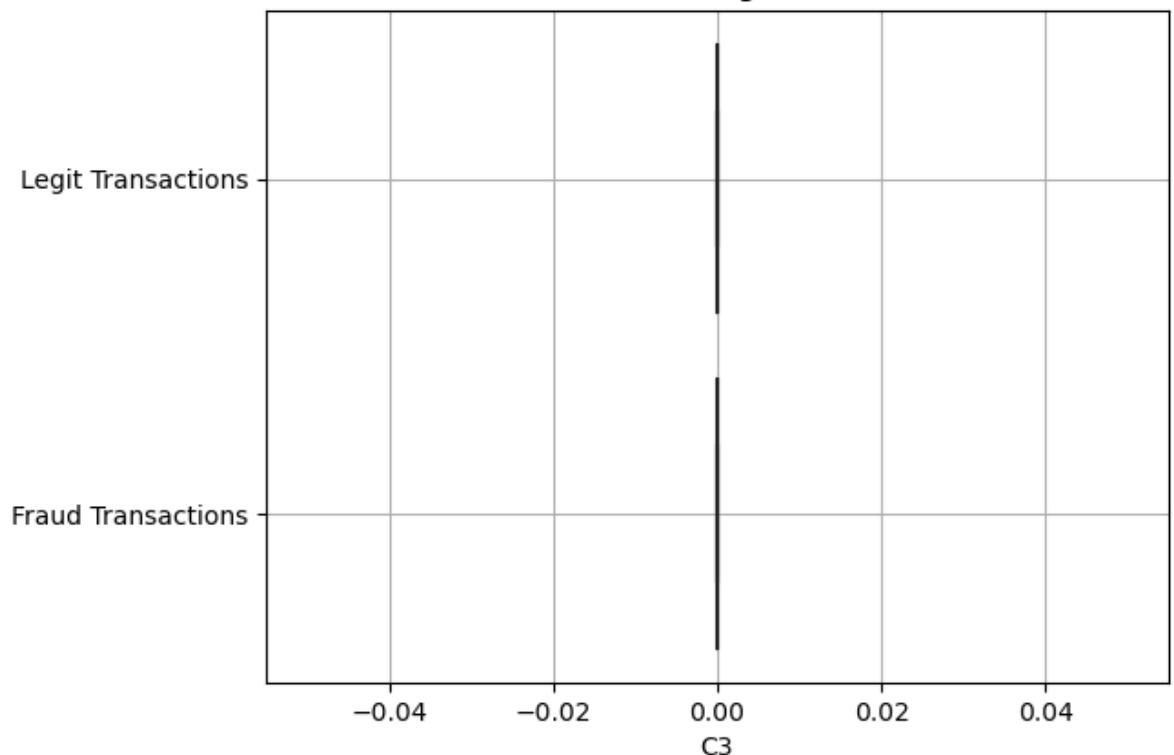
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	1.0	1.0
10	1.0	1.0
15	1.0	1.0
20	1.0	1.0
25	1.0	1.0
30	1.0	1.0
35	1.0	1.0
40	1.0	1.0
45	1.0	2.0
50	1.0	2.0
55	1.0	2.0
60	2.0	3.0
65	2.0	3.0
70	2.0	4.0
75	2.0	4.0
80	3.0	5.0
85	4.0	7.0
90	5.0	10.0
95	7.0	15.0
100	27.0	27.0

InterQuartileRange of legit transactions: 1.0

InterQuartileRange of fraud transactions: 3.0

C2 of Legit Transaction and Fraud Transaction

Analysis of C3 feature**Box Plots of C3 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	588111.0			567453.0
mean	20658.0			
std	0.0			
min	0.0			
25%	0.0			25%
50%	0.0			0.0
75%	0.0			75%
max	0.0			0.0
Name: C3, dtype: float64		Name: C3, dtype: float64		
Name: C3, dtype: float64				

The unique values in C3 feature is: 1

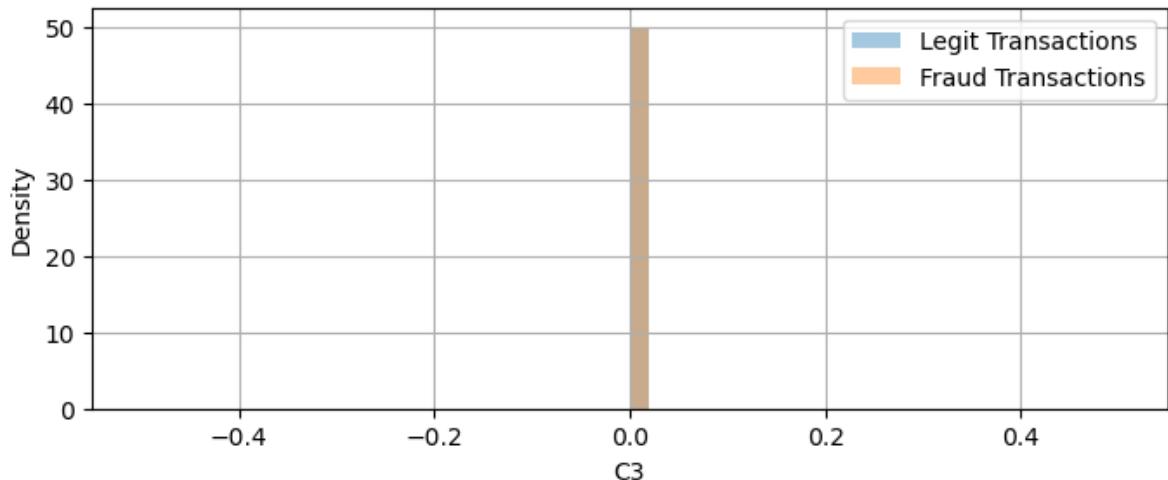
C3 feature has 0 null values, which holds 0.0 %

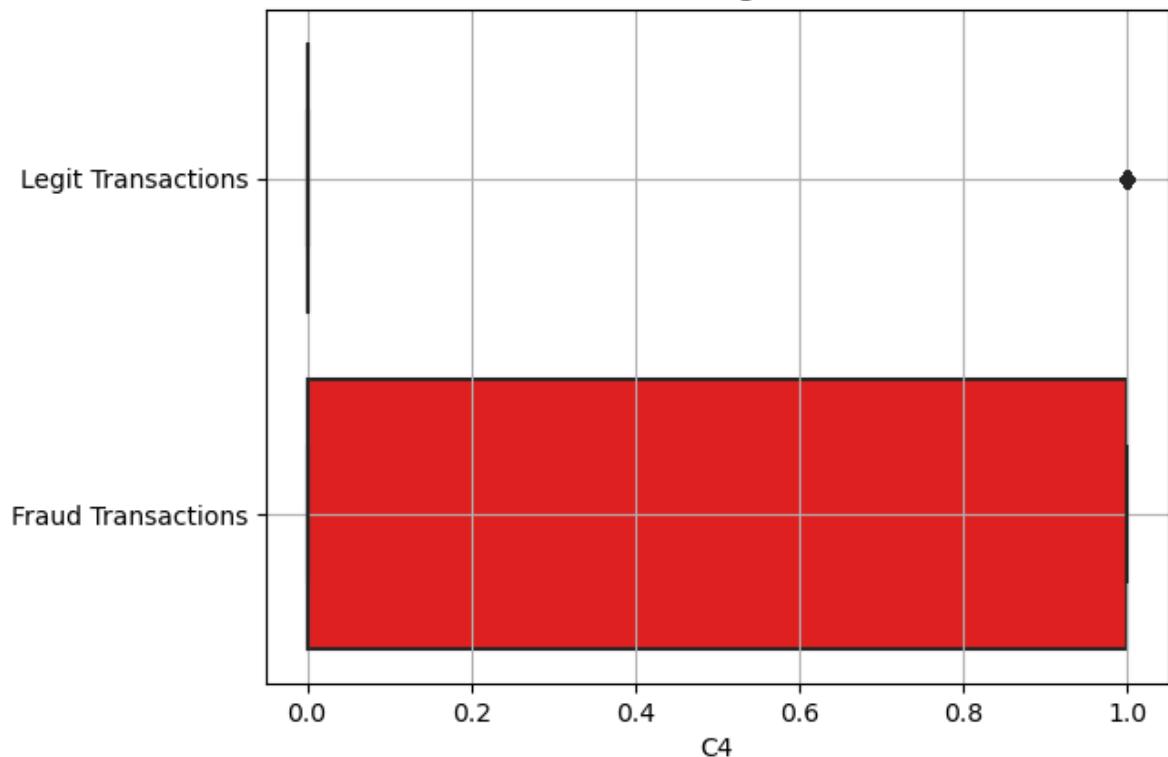
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	0.0
80	0.0	0.0
85	0.0	0.0
90	0.0	0.0
95	0.0	0.0
100	0.0	0.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 0.0

C3 of Legit Transaction and Fraud Transaction

Analysis of C4 feature**Box Plots of C4 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	563700.000000	count	548752.000000	14948.000000
mean	0.198363	mean	0.193313	0.38373
std	0.398767	std	0.394897	0.48631
min	0.000000	min	0.000000	0.000000
25%	0.000000	25%	0.000000	0.000000
50%	0.000000	50%	0.000000	0.000000
75%	0.000000	75%	0.000000	1.000000
max	1.000000	max	1.000000	1.000000
Name: C4, dtype: float64		Name: C4, dtype: float64		Name: C4, dtype: float64

The unique values in C4 feature is: 2

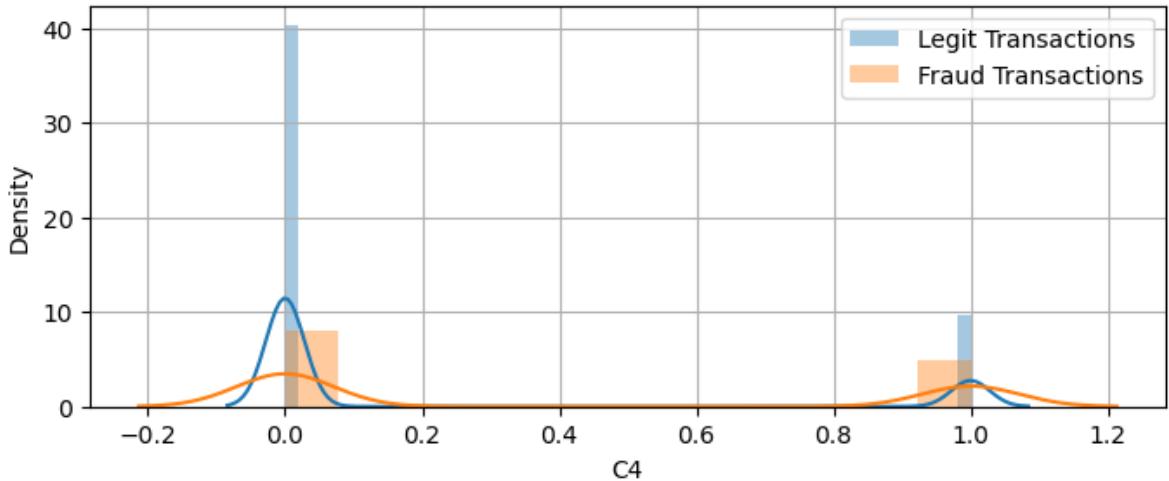
C4 feature has 0 null values, which holds 0.0 %

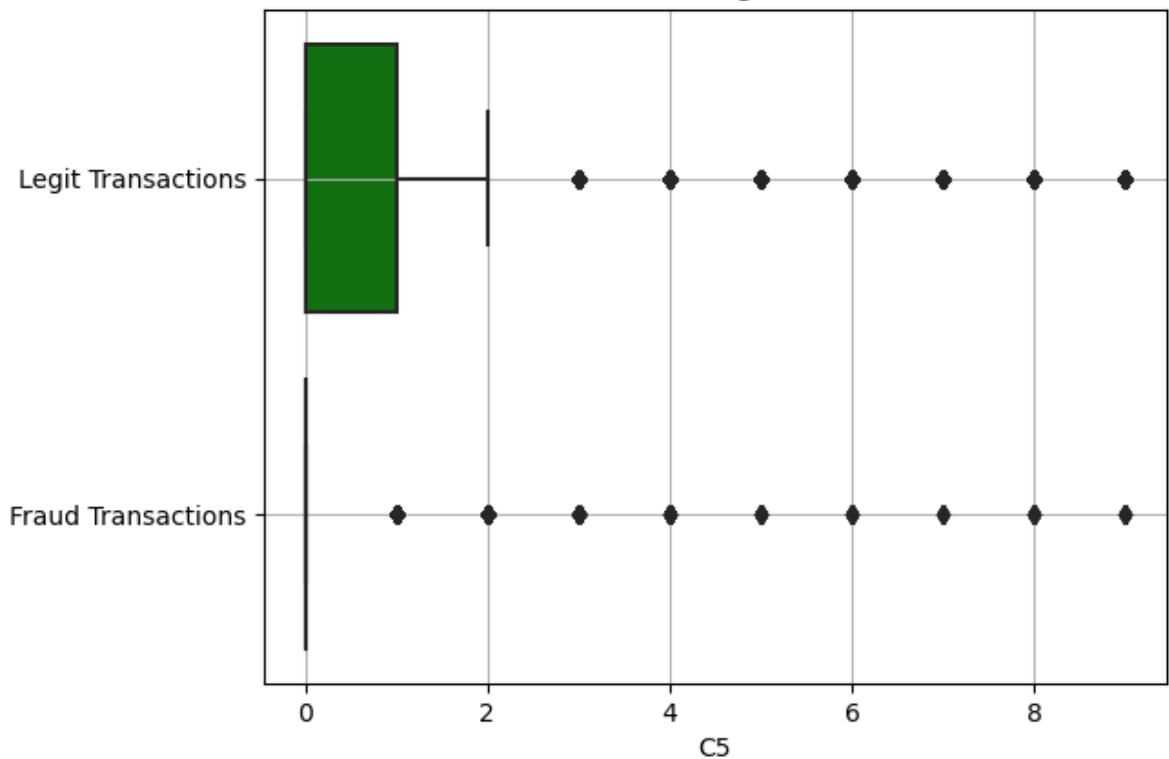
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	1.0
70	0.0	1.0
75	0.0	1.0
80	0.0	1.0
85	1.0	1.0
90	1.0	1.0
95	1.0	1.0
100	1.0	1.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 1.0

C4 of Legit Transaction and Fraud Transaction

Analysis of C5 feature**Box Plots of C5 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	561136.000000	count	540806.000000	20330.000000
mean	0.587596	mean	0.604618	0.134776
std	1.167839	std	1.180719	0.588589
min	0.000000	min	0.000000	0.000000
25%	0.000000	25%	0.000000	0.000000
50%	0.000000	50%	0.000000	0.000000
75%	1.000000	75%	1.000000	0.000000
max	9.000000	max	9.000000	9.000000
Name: C5, dtype: float64		Name: C5, dtype: float64		Name: C5, dtype: float64

The unique values in C5 feature is: 10

C5 feature has 0 null values, which holds 0.0 %

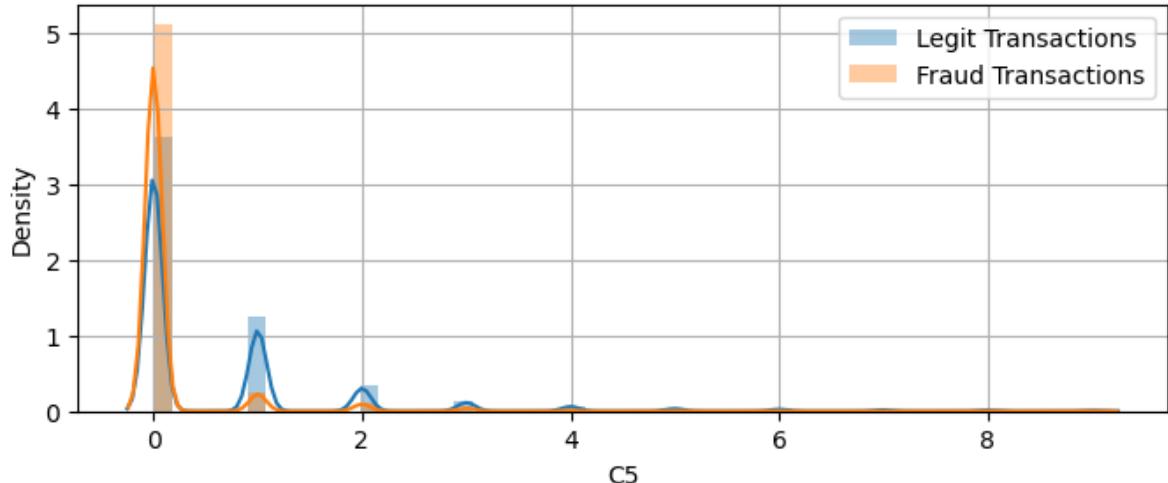
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	1.0	0.0
75	1.0	0.0
80	1.0	0.0
85	1.0	0.0
90	2.0	0.0
95	3.0	1.0
100	9.0	9.0

InterQuartileRange of legit transactions: 1.0

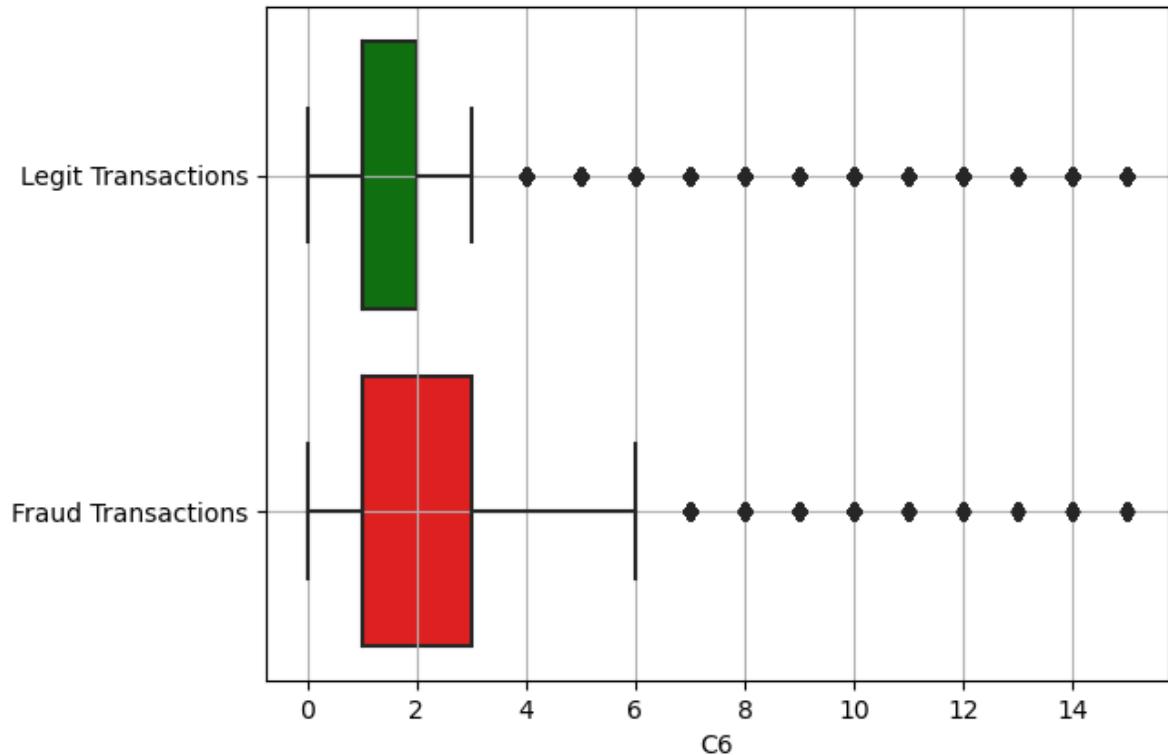
InterQuartileRange of fraud transactions: 0.0

C5 of Legit Transaction and Fraud Transaction



Analysis of C6 feature

Box Plots of C6 based on legit and fraud trasactions



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	561481.000000	count	542375.000000	19106.000000
mean	1.641122	mean	1.619726	2.248508
std	1.724284	std	1.686174	2.506230
min	0.000000	min	0.000000	0.000000
25%	1.000000	25%	1.000000	1.000000
50%	1.000000	50%	1.000000	1.000000
75%	2.000000	75%	2.000000	3.000000
max	15.000000	max	15.000000	15.000000
Name: C6, dtype: float64		Name: C6, dtype: float64		Name: C6, dtype: float64

The unique values in C6 feature is: 16

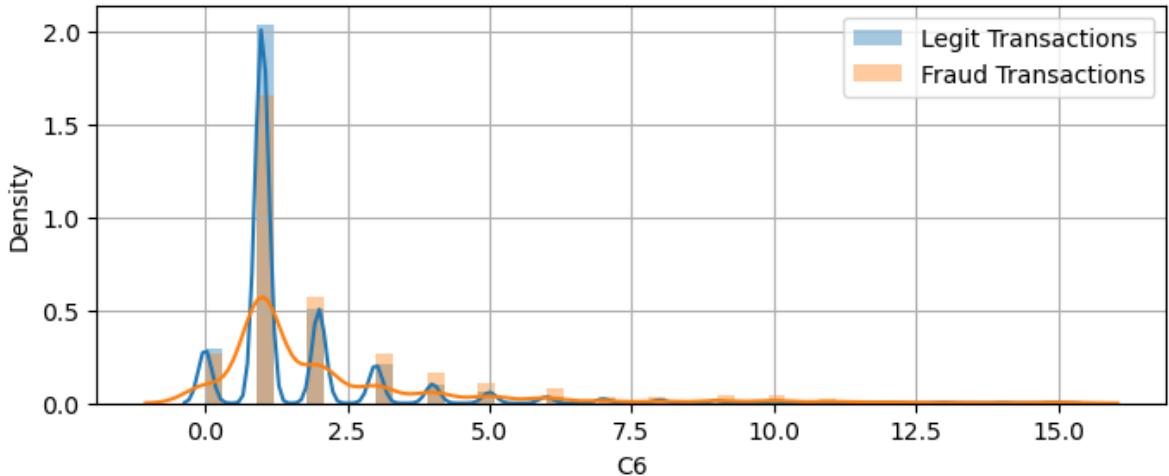
C6 feature has 0 null values, which holds 0.0 %

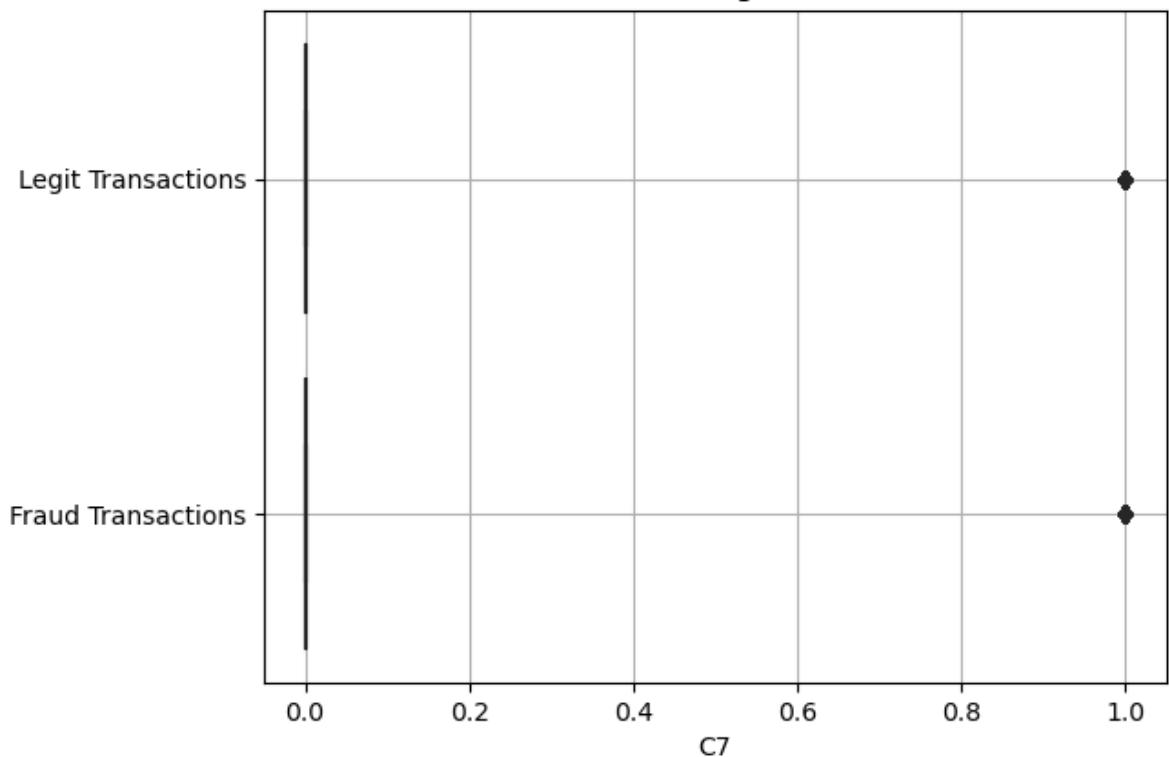
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	1.0	1.0
15	1.0	1.0
20	1.0	1.0
25	1.0	1.0
30	1.0	1.0
35	1.0	1.0
40	1.0	1.0
45	1.0	1.0
50	1.0	1.0
55	1.0	1.0
60	1.0	2.0
65	1.0	2.0
70	2.0	2.0
75	2.0	3.0
80	2.0	3.0
85	2.0	4.0
90	3.0	5.0
95	5.0	8.0
100	15.0	15.0

InterQuartileRange of legit transactions: 1.0

InterQuartileRange of fraud transactions: 2.0

C6 of Legit Transaction and Fraud Transaction

Analysis of C7 feature**Box Plots of C7 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	575108.000000	count	558859.000000	16249.000000
mean	0.090359	mean	0.086728	0.215213
std	0.286695	std	0.281437	0.410983
min	0.000000	min	0.000000	0.000000
25%	0.000000	25%	0.000000	0.000000
50%	0.000000	50%	0.000000	0.000000
75%	0.000000	75%	0.000000	0.000000
max	1.000000	max	1.000000	1.000000
Name: C7, dtype: float64		Name: C7, dtype: float64		Name: C7, dtype: float64

The unique values in C7 feature is: 2

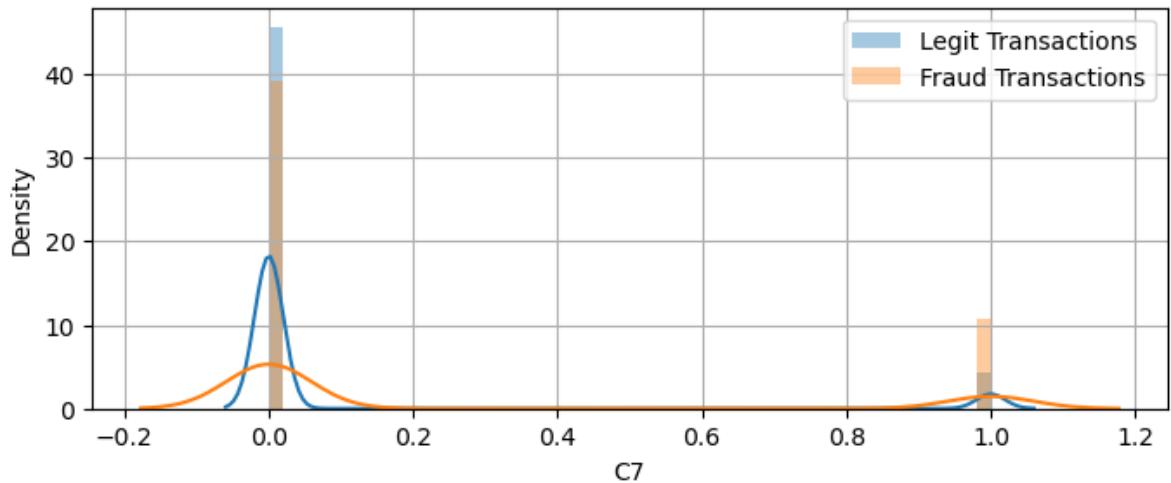
C7 feature has 0 null values, which holds 0.0 %

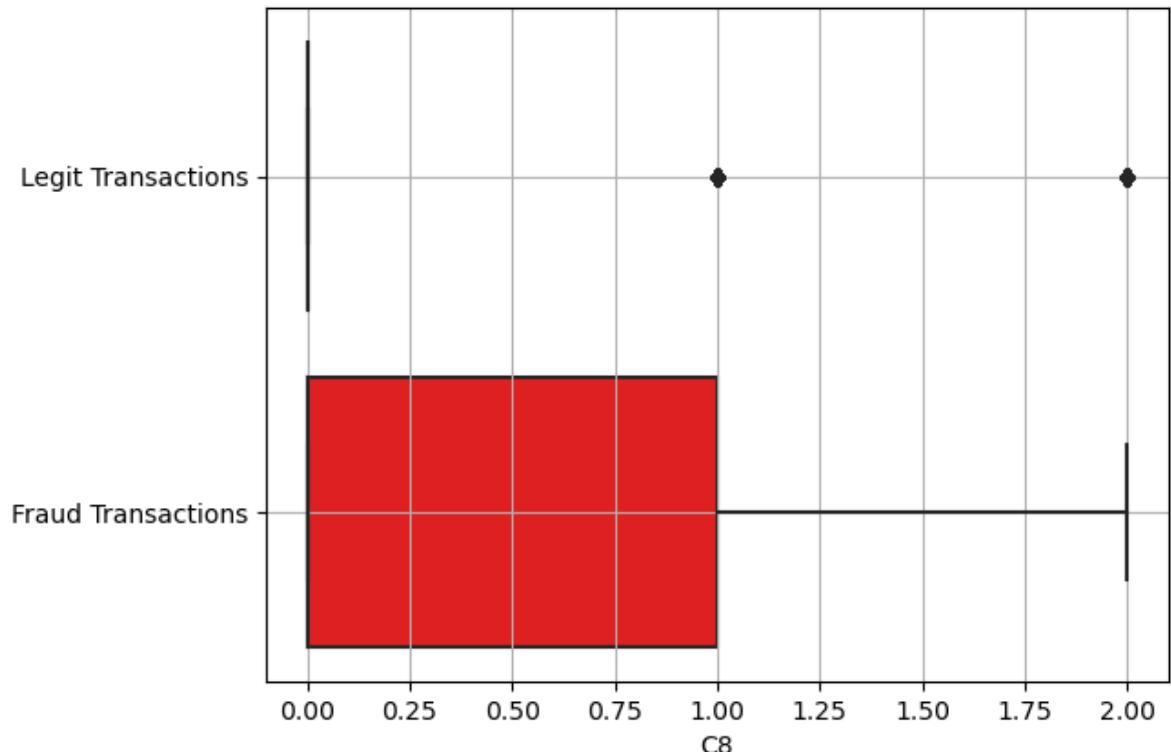
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	0.0
80	0.0	1.0
85	0.0	1.0
90	0.0	1.0
95	1.0	1.0
100	1.0	1.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 0.0

C7 of Legit Transaction and Fraud Transaction

Analysis of C8 feature**Box Plots of C8 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	568588.000000	count	553016.000000	15572.000000
mean	0.240680	mean	0.233406	0.499037
std	0.488648	std	0.480525	0.670631
min	0.000000	min	0.000000	0.000000
25%	0.000000	25%	0.000000	0.000000
50%	0.000000	50%	0.000000	0.000000
75%	0.000000	75%	0.000000	1.000000
max	2.000000	max	2.000000	2.000000
Name: C8, dtype: float64		Name: C8, dtype: float64		Name: C8, dtype: float64

The unique values in C8 feature is: 3

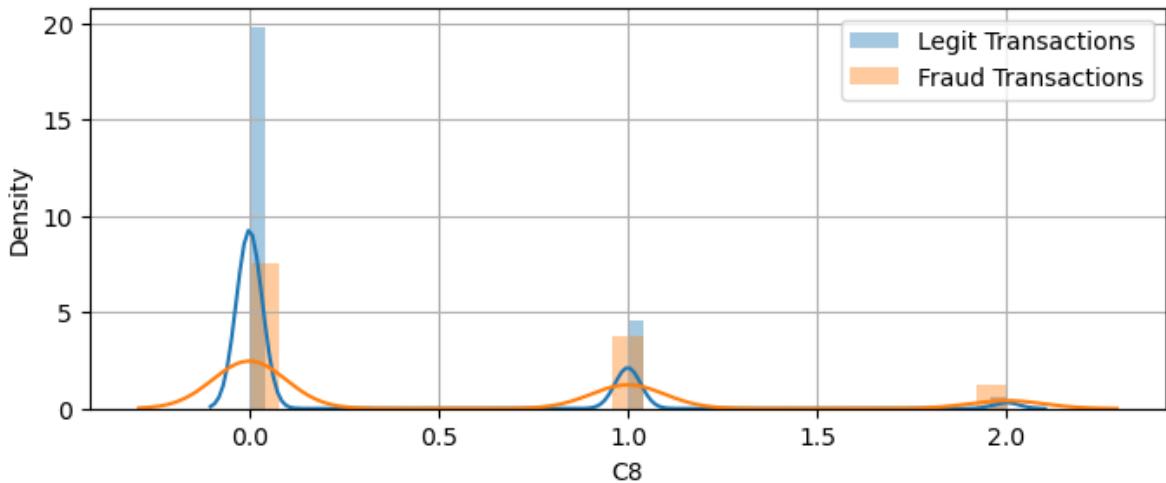
C8 feature has 0 null values, which holds 0.0 %

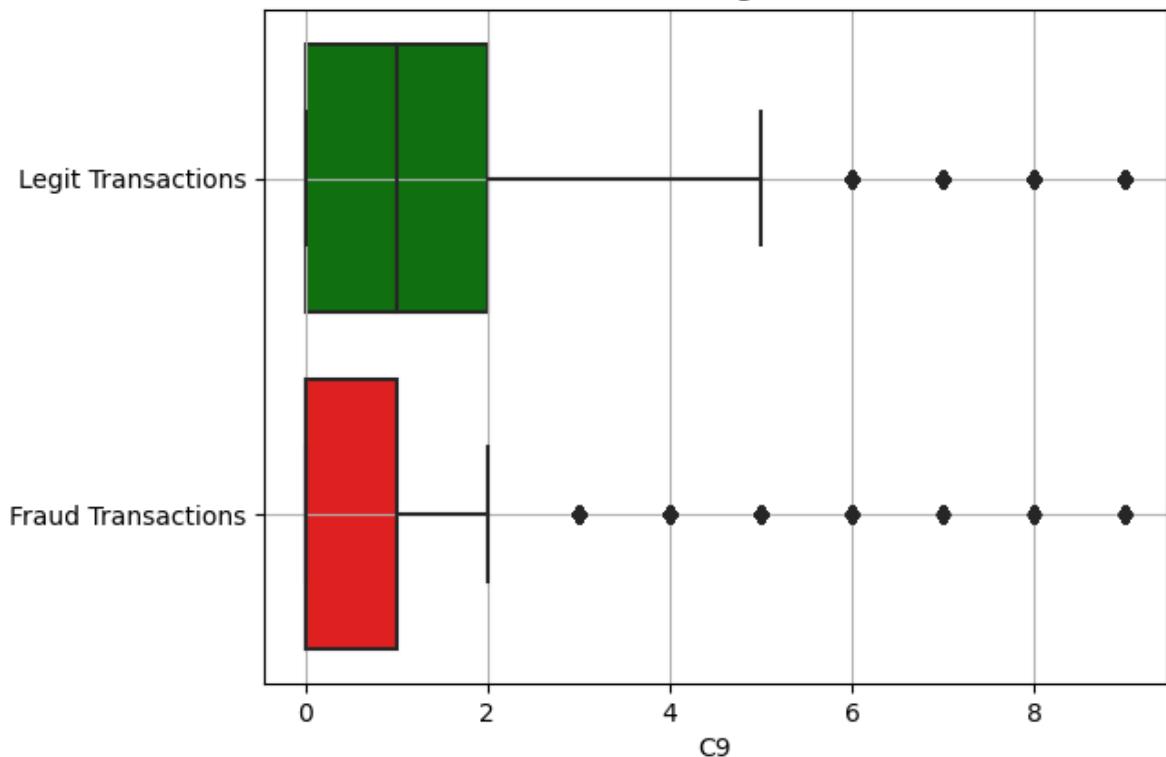
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	1.0
70	0.0	1.0
75	0.0	1.0
80	1.0	1.0
85	1.0	1.0
90	1.0	1.0
95	1.0	2.0
100	2.0	2.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 1.0

C8 of Legit Transaction and Fraud Transaction

Analysis of C9 feature**Box Plots of C9 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	561619.000000	count	541353.000000	20266.000000
mean	1.222574	mean	1.237529	0.823103
std	1.439396	std	1.435225	1.492192
min	0.000000	min	0.000000	0.000000
25%	0.000000	25%	0.000000	0.000000
50%	1.000000	50%	1.000000	0.000000
75%	2.000000	75%	2.000000	1.000000
max	9.000000	max	9.000000	9.000000
Name: C9, dtype: float64		Name: C9, dtype: float64		Name: C9, dtype: float64

The unique values in C9 feature is: 10

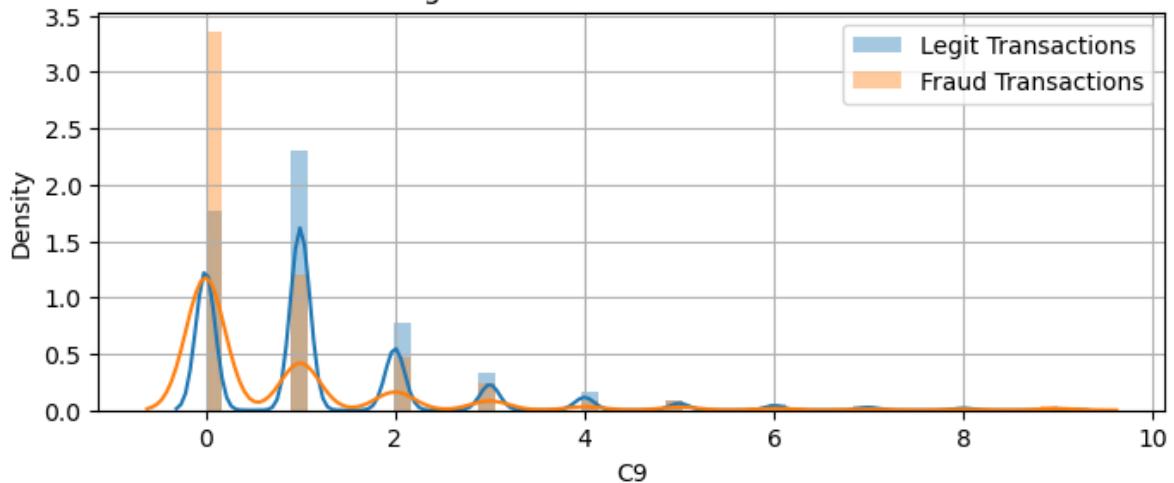
C9 feature has 0 null values, which holds 0.0 %

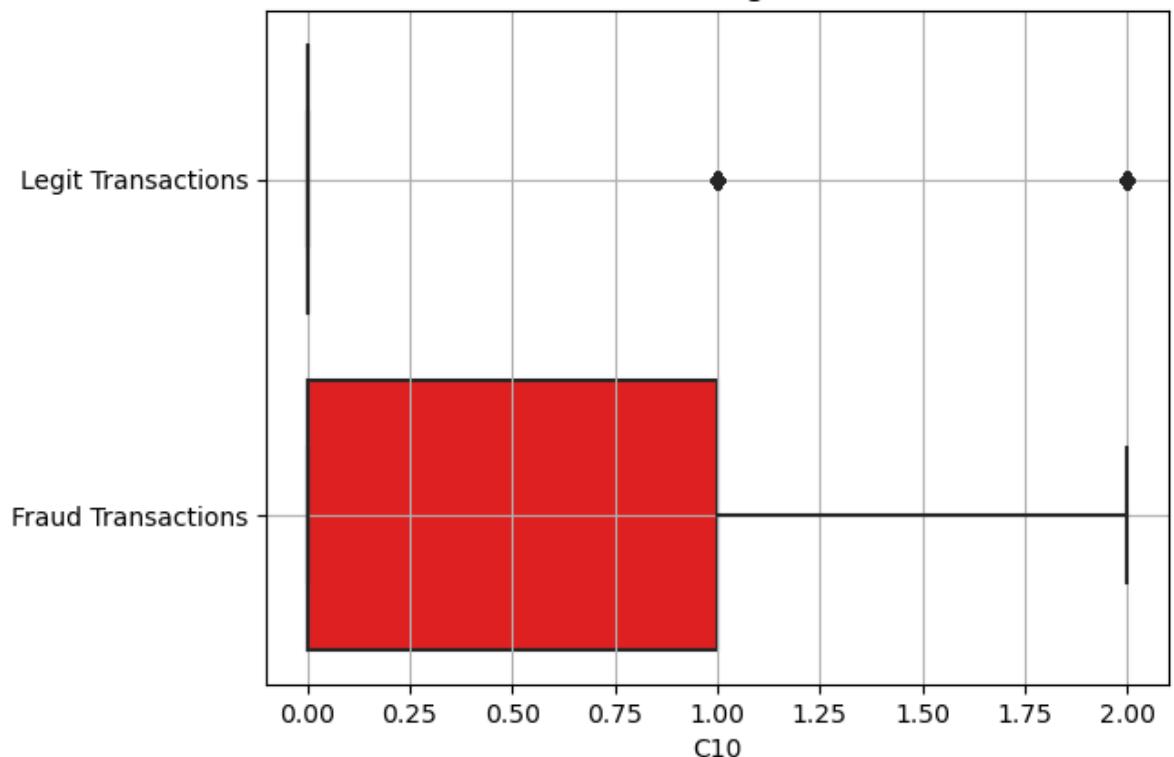
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	1.0	0.0
40	1.0	0.0
45	1.0	0.0
50	1.0	0.0
55	1.0	0.0
60	1.0	0.0
65	1.0	1.0
70	1.0	1.0
75	2.0	1.0
80	2.0	1.0
85	2.0	2.0
90	3.0	2.0
95	4.0	4.0
100	9.0	9.0

InterQuartileRange of legit transactions: 2.0

InterQuartileRange of fraud transactions: 1.0

C9 of Legit Transaction and Fraud Transaction

Analysis of C10 feature**Box Plots of C10 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	568611.000000	count	552689.000000	
count	15922.000000			
mean	0.228815	mean	0.221094	
mean	0.49686			
std	0.478541	std	0.469415	
std	0.67483			
min	0.000000	min	0.000000	
min	0.00000			
25%	0.000000	25%	0.000000	
25%	0.00000			
50%	0.000000	50%	0.000000	
50%	0.00000			
75%	0.000000	75%	0.000000	
75%	1.00000			
max	2.000000	max	2.000000	
max	2.00000			
Name: C10, dtype: float64		Name: C10, dtype: float64		
Name: C10, dtype: float64				

The unique values in C10 feature is: 3

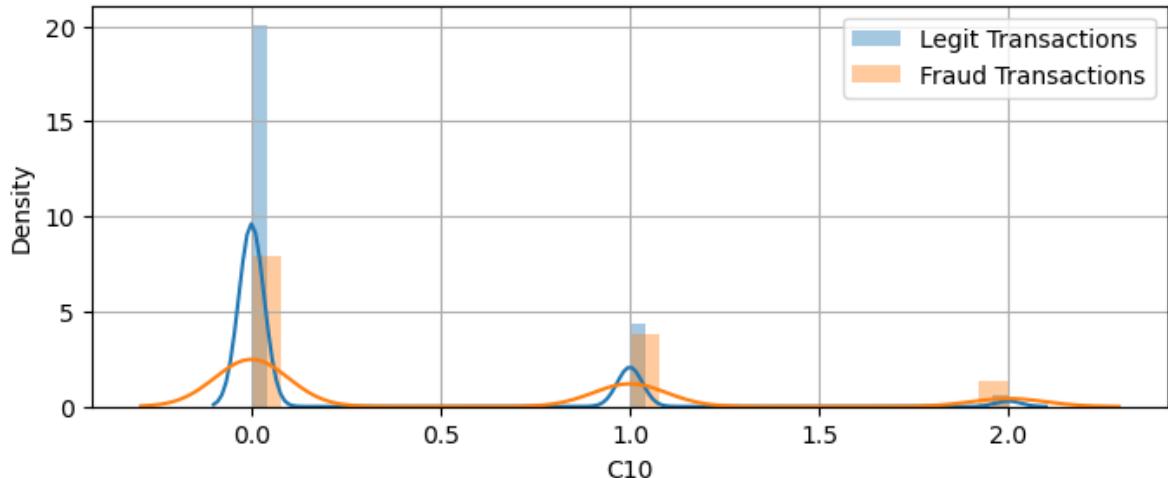
C10 feature has 0 null values, which holds 0.0 %

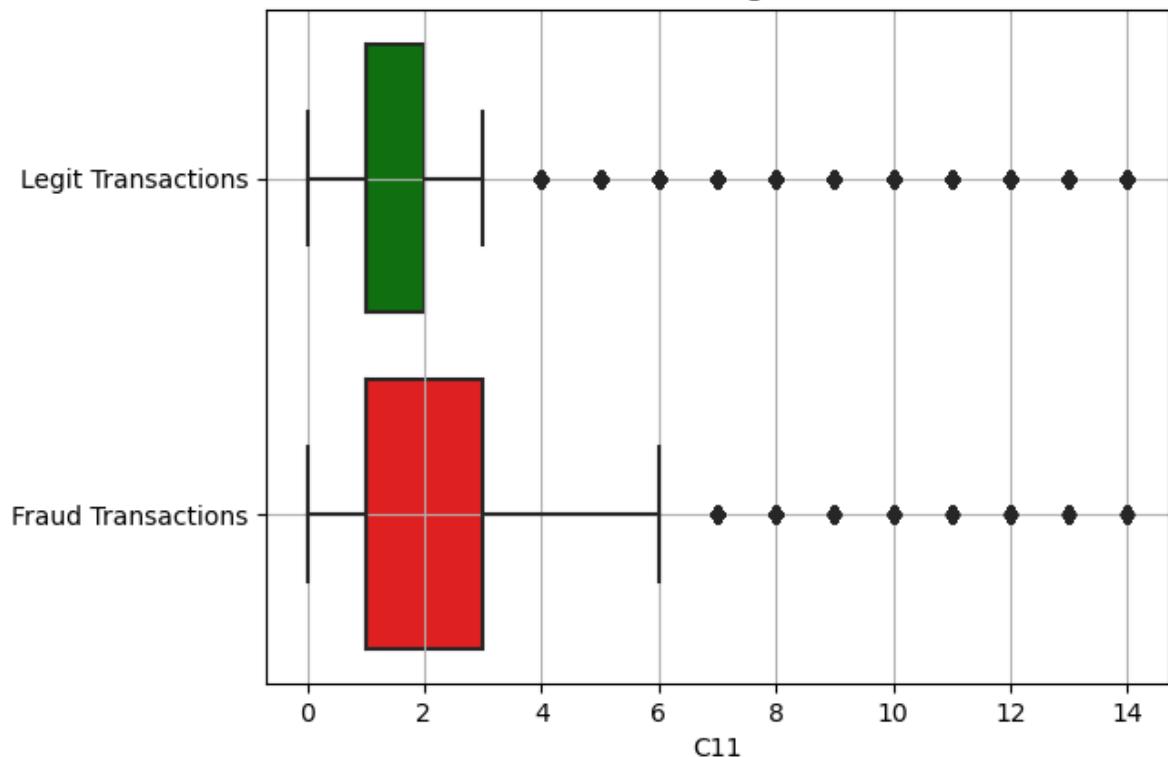
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	1.0
70	0.0	1.0
75	0.0	1.0
80	0.0	1.0
85	1.0	1.0
90	1.0	2.0
95	1.0	2.0
100	2.0	2.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 1.0

C10 of Legit Transaction and Fraud Transaction

Analysis of C11 feature**Box Plots of C11 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	561018.000000	count	542573.000000	
count	18445.000000			
mean	1.683534	mean	1.662010	
mean	2.316671			
std	1.560959	std	1.523695	
std	2.324580			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	2.000000	75%	2.000000	
75%	3.000000			
max	14.000000	max	14.000000	
max	14.000000			
Name: C11, dtype: float64		Name: C11, dtype: float64		
Name: C11, dtype: float64				

The unique values in C11 feature is: 15

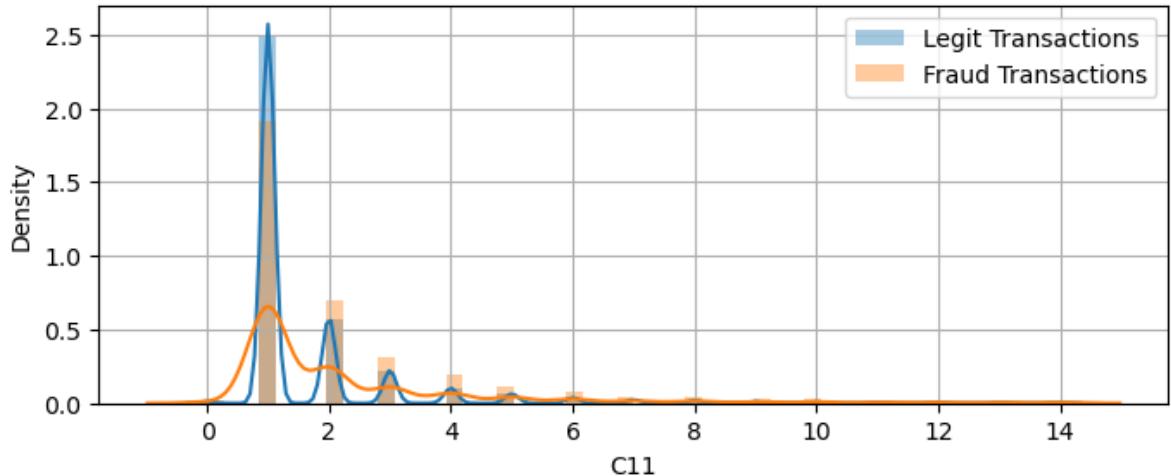
C11 feature has 0 null values, which holds 0.0 %

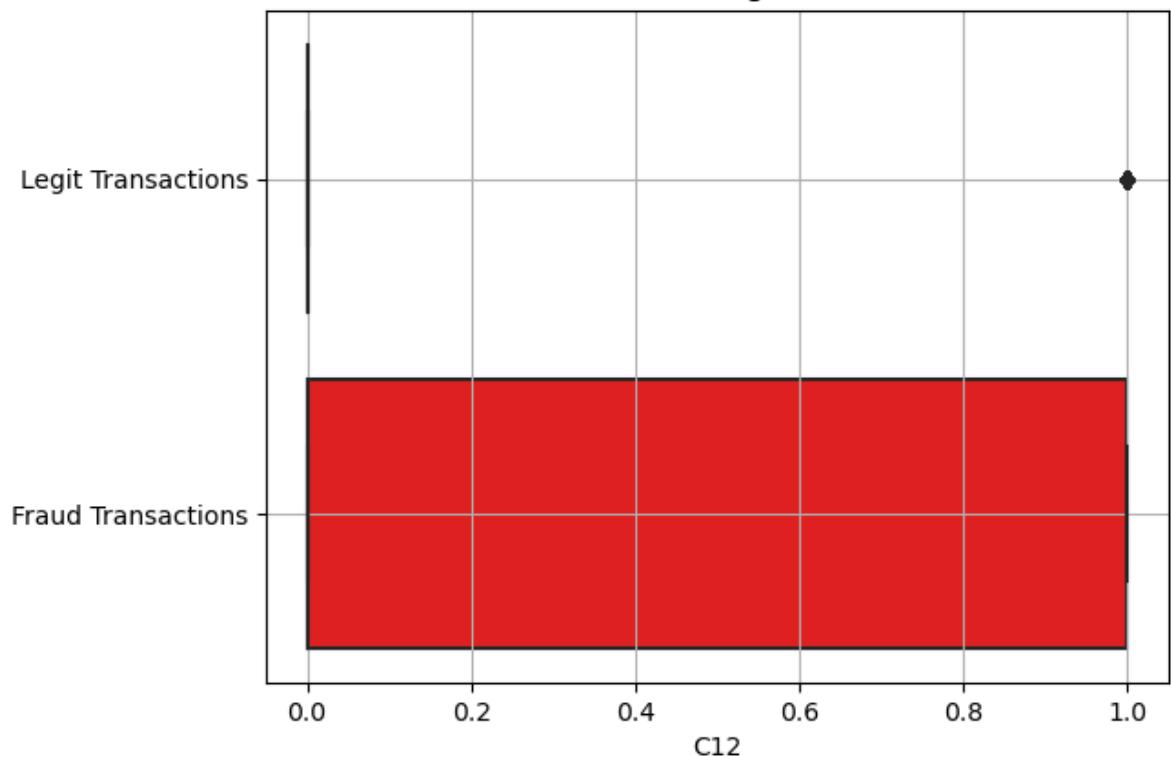
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	1.0	1.0
10	1.0	1.0
15	1.0	1.0
20	1.0	1.0
25	1.0	1.0
30	1.0	1.0
35	1.0	1.0
40	1.0	1.0
45	1.0	1.0
50	1.0	1.0
55	1.0	2.0
60	1.0	2.0
65	1.0	2.0
70	1.0	2.0
75	2.0	3.0
80	2.0	3.0
85	2.0	4.0
90	3.0	5.0
95	4.0	7.0
100	14.0	14.0

InterQuartileRange of legit transactions: 1.0

InterQuartileRange of fraud transactions: 2.0

C11 of Legit Transaction and Fraud Transaction

Analysis of C12 feature**Box Plots of C12 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	565860.000000	count	550570.000000	
count	15290.000000			
mean	0.135477	mean	0.131812	
mean	0.267430			
std	0.342233	std	0.338287	
std	0.442633			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	1.000000	max	1.000000	
max	1.000000			
Name: C12, dtype: float64		Name: C12, dtype: float64		
Name: C12, dtype: float64				

The unique values in C12 feature is: 2

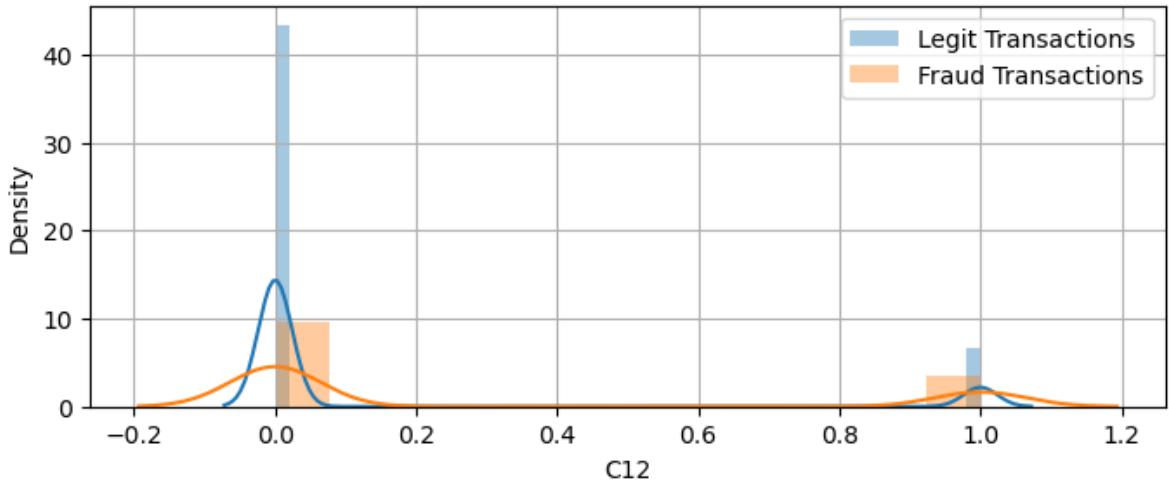
C12 feature has 0 null values, which holds 0.0 %

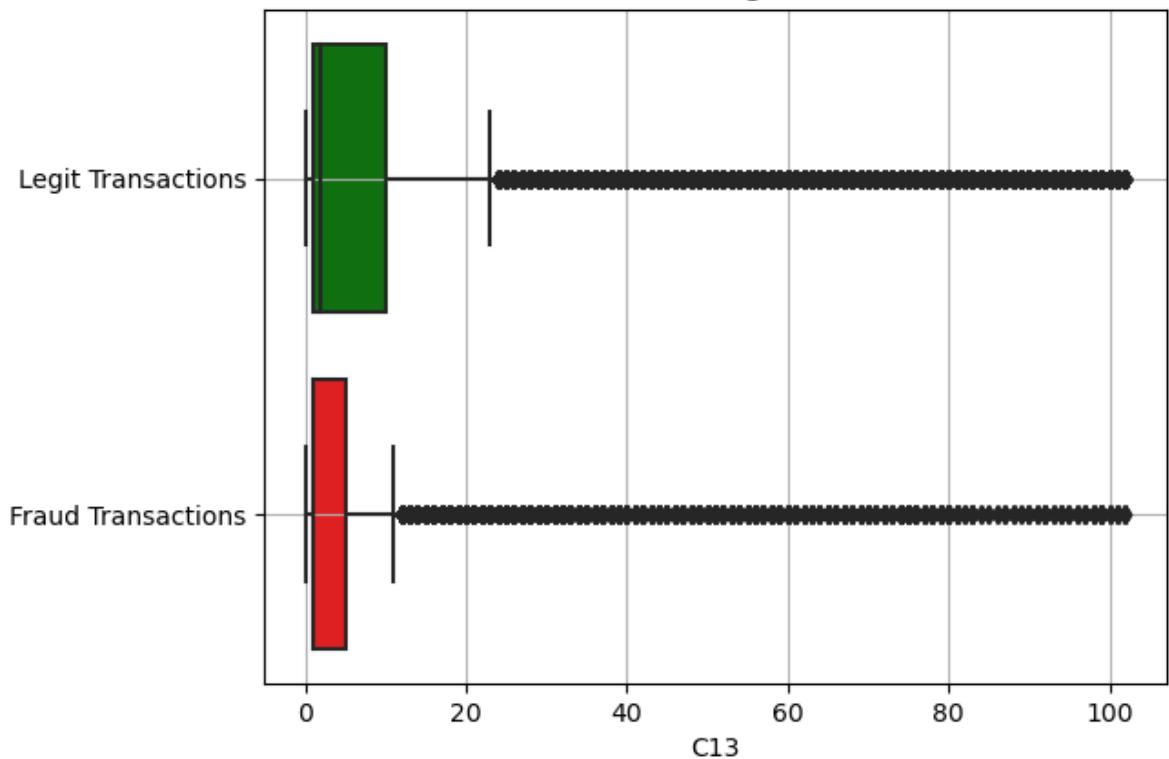
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	1.0
80	0.0	1.0
85	0.0	1.0
90	1.0	1.0
95	1.0	1.0
100	1.0	1.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 1.0

C12 of Legit Transaction and Fraud Transaction

Analysis of C13 feature**Box Plots of C13 based on legit and fraud transactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	561143.000000	count	540988.000000	
count	20155.000000			
mean	8.614854	mean	8.713101	
mean	5.977772			
std	14.306454	std	14.355356	
std	12.643119			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	2.000000	50%	2.000000	
50%	1.000000			
75%	10.000000	75%	10.000000	
75%	5.000000			
max	102.000000	max	102.000000	
max	102.000000			
Name: C13, dtype: float64		Name: C13, dtype: float64		
Name: C13, dtype: float64				

The unique values in C13 feature is: 103

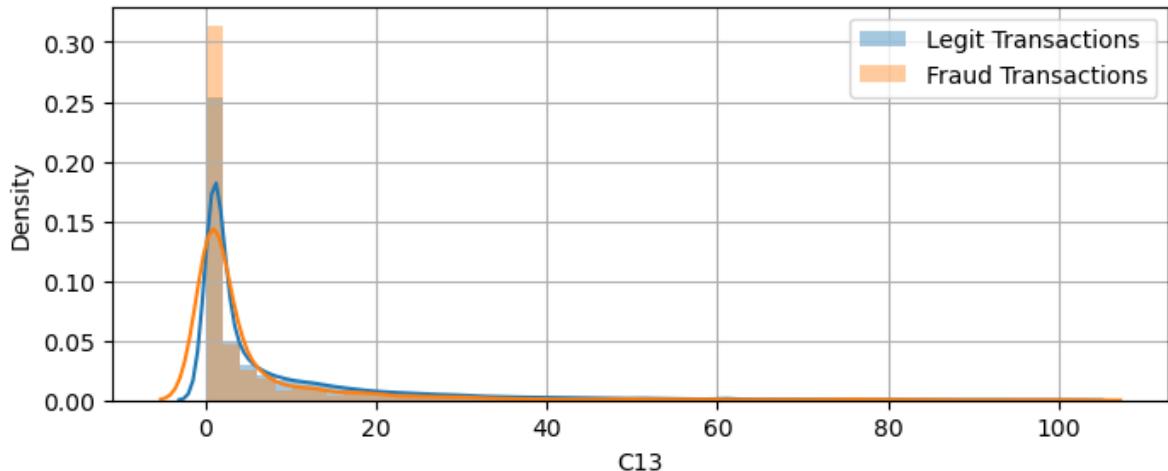
C13 feature has 0 null values, which holds 0.0 %

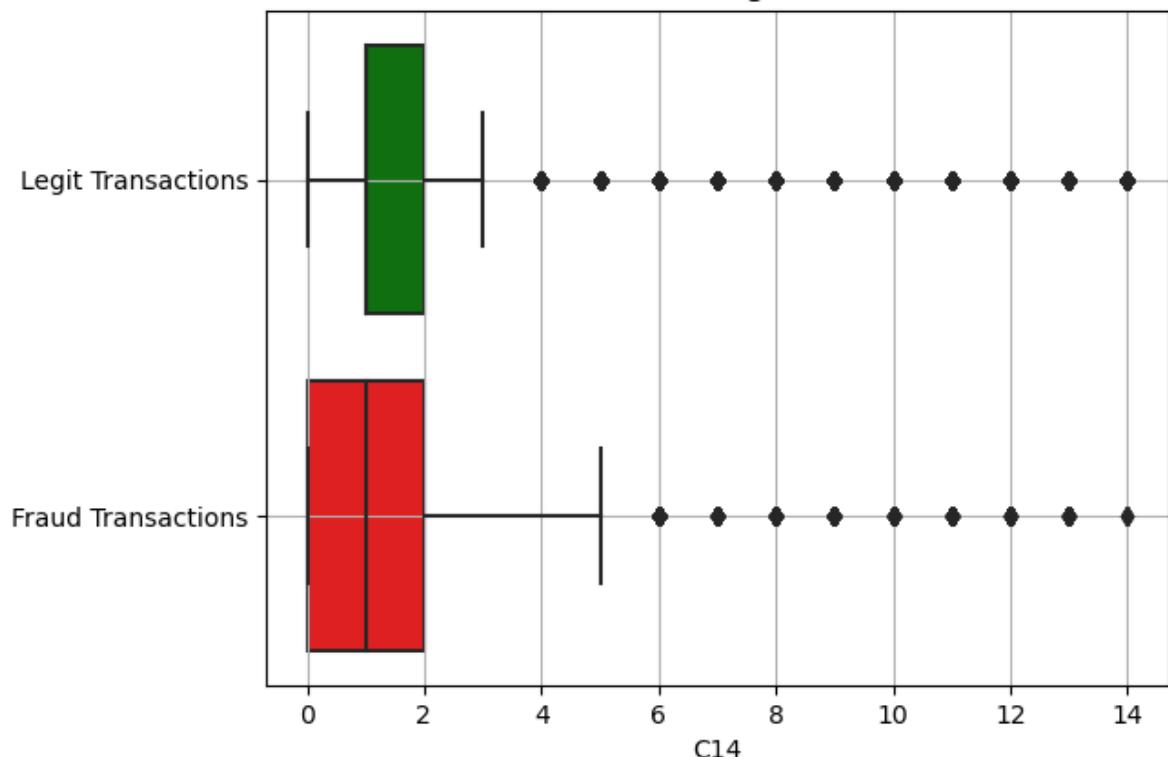
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	1.0	0.0
15	1.0	0.0
20	1.0	0.0
25	1.0	1.0
30	1.0	1.0
35	1.0	1.0
40	1.0	1.0
45	2.0	1.0
50	2.0	1.0
55	3.0	2.0
60	4.0	2.0
65	5.0	3.0
70	7.0	4.0
75	10.0	5.0
80	13.0	7.0
85	18.0	11.0
90	25.0	17.0
95	39.0	29.0
100	102.0	102.0

InterQuartileRange of legit transactions: 9.0

InterQuartileRange of fraud transactions: 4.0

C13 of Legit Transaction and Fraud Transaction

Analysis of C14 feature**Box Plots of C14 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	561456.000000	count	541393.000000	
mean	1.835088	mean	1.840607	
std	1.686139	std	1.776471	
min	0.000000	min	0.000000	
25%	0.000000	25%	1.000000	
50%	1.000000	50%	1.000000	
75%	2.000000	75%	2.000000	
max	14.000000	max	14.000000	
	Name: C14, dtype: float64		Name: C14, dtype: float64	
	Name: C14, dtype: float64			

The unique values in C14 feature is: 15

C14 feature has 0 null values, which holds 0.0 %

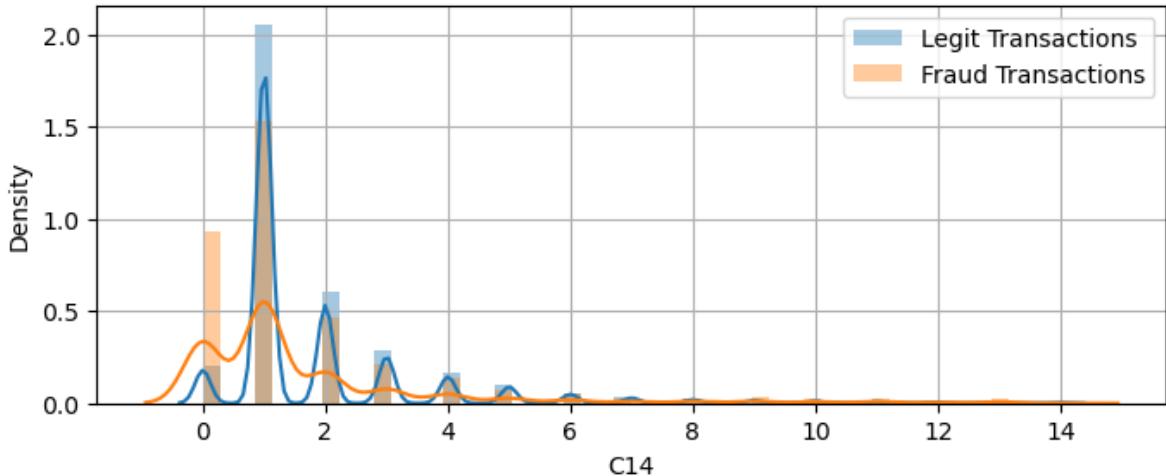
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	1.0	0.0
15	1.0	0.0
20	1.0	0.0
25	1.0	0.0
30	1.0	1.0
35	1.0	1.0
40	1.0	1.0
45	1.0	1.0
50	1.0	1.0
55	1.0	1.0
60	1.0	1.0
65	2.0	1.0
70	2.0	2.0
75	2.0	2.0
80	2.0	2.0
85	3.0	3.0
90	4.0	4.0
95	5.0	6.0
100	14.0	14.0

InterQuartileRange of legit transactions: 1.0

InterQuartileRange of fraud transactions: 2.0

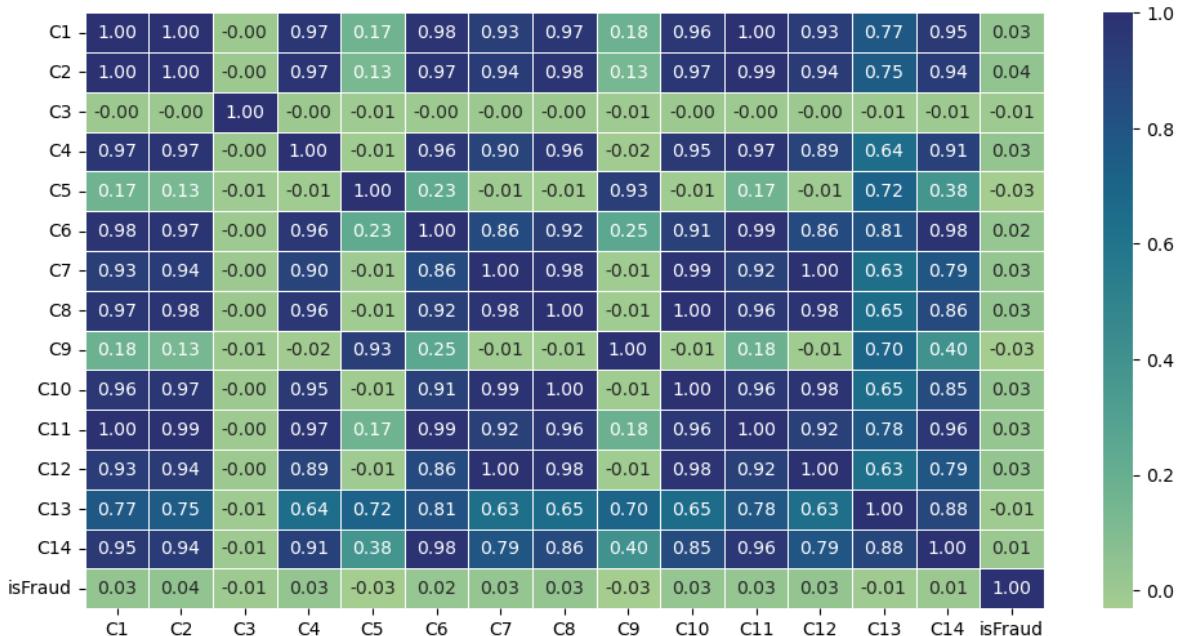
C14 of Legit Transaction and Fraud Transaction



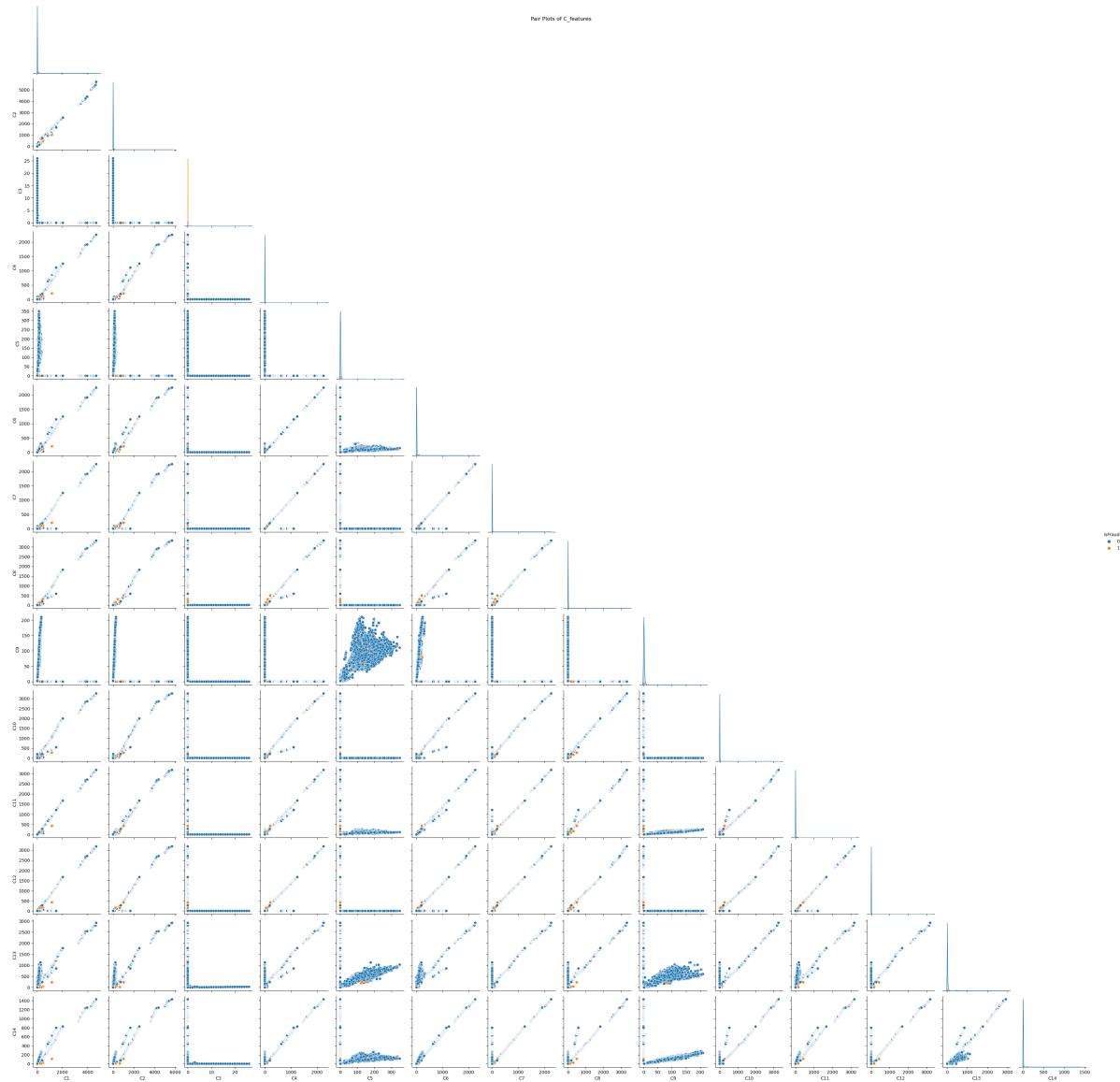
```
*****
*****
```

In [74]: # to get collinearity of C features

```
plt.figure(figsize=(12,6))
sns.heatmap(df_train[c_features + ['isFraud']].corr(), annot = True, fmt=".2f", cmap='viridis')
plt.show()
```



In [75]: sns.pairplot(vars = c_features, hue='isFraud', data = df_train, corner=True)
plt.suptitle("\nPair Plots of C_features\n")
plt.show()



Observation:

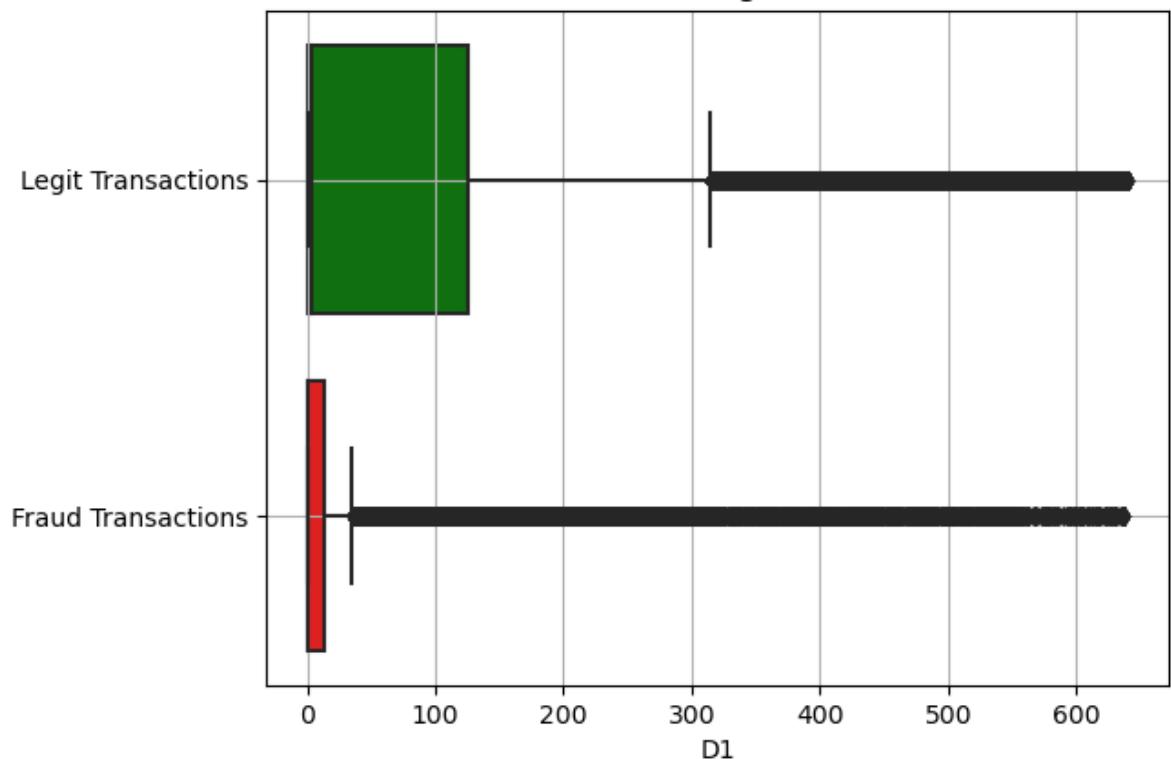
1. These features are extremely correlated, and it is therefore necessary to remove them.
2. The distribution of fraudulent transactions is much more broadly dispersed than that of legitimate transactions for feature C1, C2, C4, C6, C7, C8, C10, C11, C12, and C14. It appears that as the value of these characteristics rises, the likelihood of a fraudulent transaction increasing.
3. If the value increases, the likelihood that it is a legitimate transaction rather than a fraudulent one is increased by features C5, C9, and C13.
4. Because the C3 feature is a zero variance feature, it is useless, so we will eliminate it going forward.

1.2.5 Feature: 'D features'

```
In [76]: d_features = ['D1', 'D2', 'D3', 'D4', 'D5', 'D6', 'D7', 'D8', 'D9', 'D10', 'D11',  
for fea in d_features:  
    print(f'Analysis of {fea} feature')  
    analysis_num_fea(df_train, fea, 'isFraud')  
    print('\n')  
    print('*'*120)  
    print('\n')
```

Analysis of D1 feature

Box Plots of D1 based on legit and fraud trasactions



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	589271.000000	count	568654.000000	
mean	94.347568	mean	96.364705	
std	157.660387	std	158.973258	
min	0.000000	min	0.000000	
25%	0.000000	25%	0.000000	
50%	3.000000	50%	4.000000	
75%	122.000000	75%	126.000000	
max	640.000000	max	640.000000	
Name: D1, dtype: float64		Name: D1, dtype: float64		
Name: D1, dtype: float64				

The unique values in D1 feature is: 642

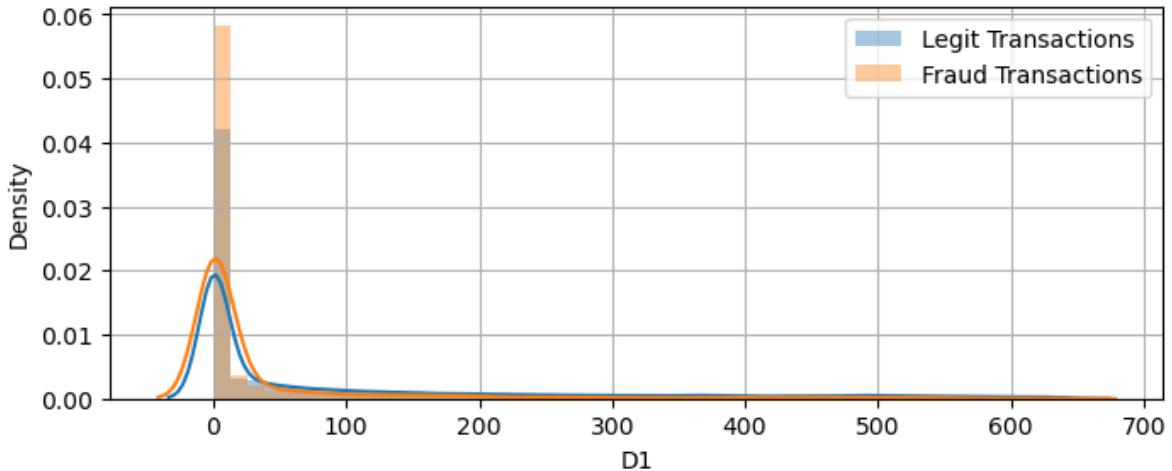
D1 feature has 1269 null values, which holds 0.0021488806854743116 %

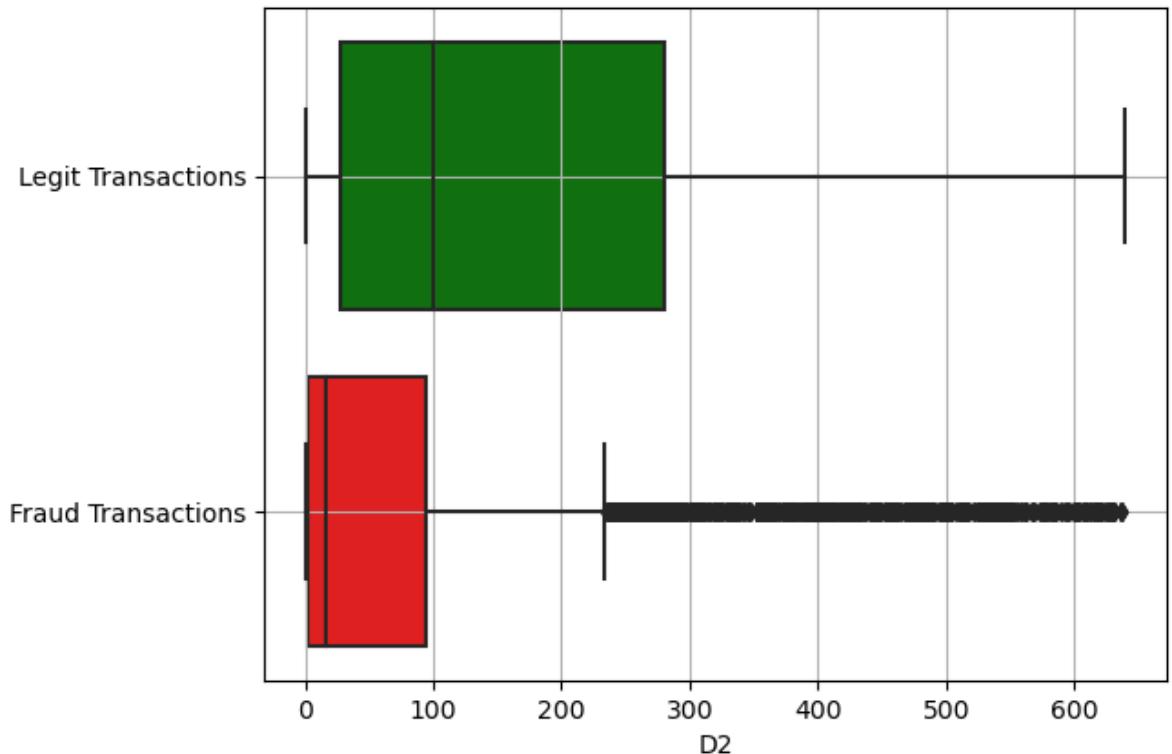
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	4.0	0.0
55	15.0	0.0
60	31.0	1.0
65	55.0	3.0
70	85.0	6.0
75	126.0	14.0
80	184.0	28.0
85	264.0	63.0
90	367.0	129.0
95	491.0	254.0
100	640.0	637.0

InterQuartileRange of legit transactions: 126.0

InterQuartileRange of fraud transactions: 14.0

D1 of Legit Transaction and Fraud Transaction

Analysis of D2 feature**Box Plots of D2 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	309743.000000	count	301873.000000	7870.000000
mean	169.563231	mean	171.956223	77.774206
std	177.315865	std	177.773639	128.697847
min	0.000000	min	0.000000	0.000000
25%	26.000000	25%	27.000000	2.000000
50%	97.000000	50%	100.000000	16.000000
75%	276.000000	75%	280.000000	95.000000
max	640.000000	max	640.000000	637.000000
Name: D2, dtype: float64		Name: D2, dtype: float64		Name: D2, dtype: float64

The unique values in D2 feature is: 642

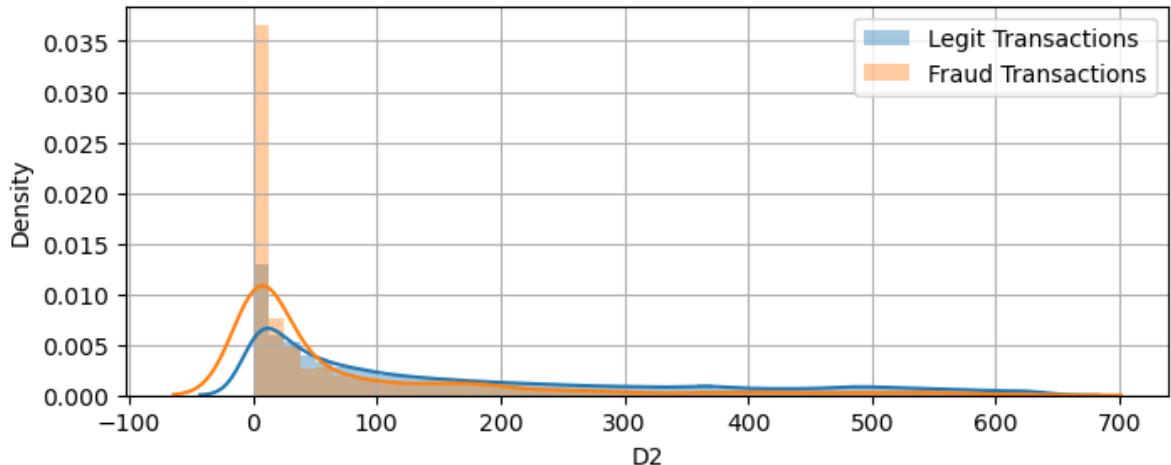
D2 feature has 280797 null values, which holds 0.4754919226470688 %

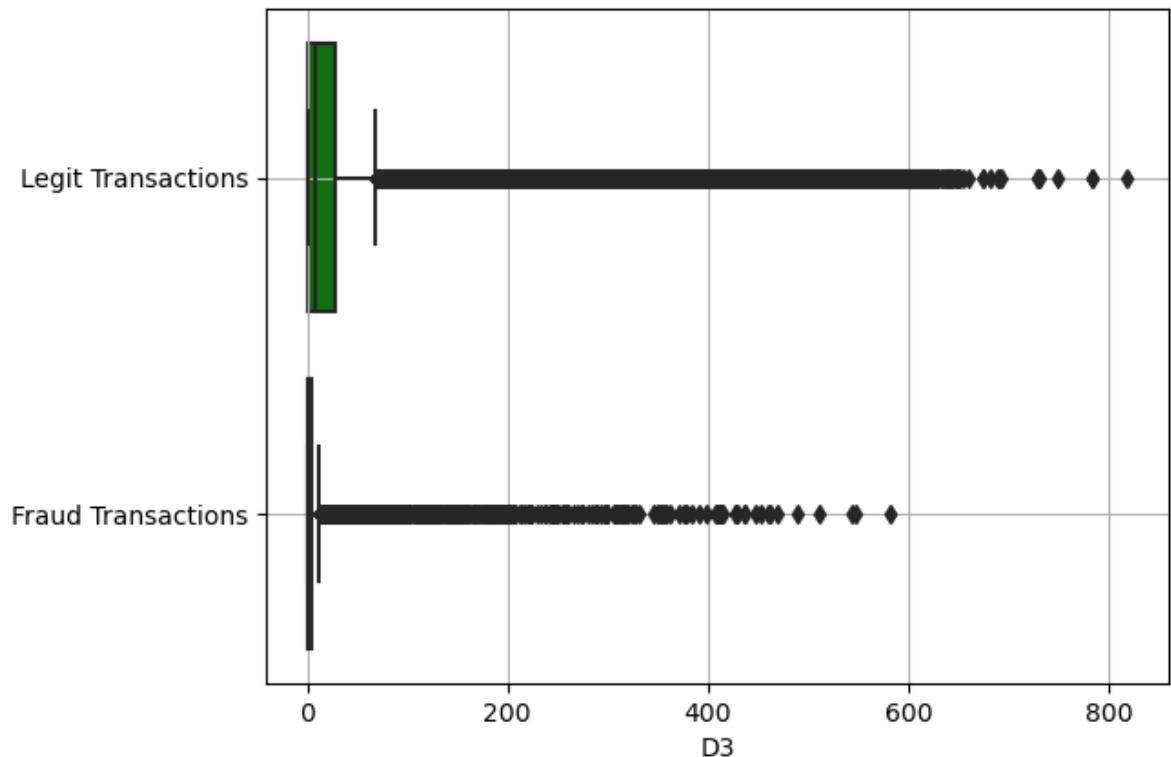
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	1.0	0.0
10	4.0	0.0
15	10.0	1.0
20	17.0	1.0
25	27.0	2.0
30	36.0	3.0
35	48.0	5.0
40	63.0	7.0
45	81.0	11.0
50	100.0	16.0
55	125.0	24.0
60	154.0	32.0
65	189.0	46.85
70	231.0	65.0
75	280.0	95.0
80	340.0	139.0
85	402.0	179.0
90	478.0	245.0
95	541.0	401.0
100	640.0	637.0

InterQuartileRange of legit transactions: 253.0

InterQuartileRange of fraud transactions: 93.0

D2 of Legit Transaction and Fraud Transaction

Analysis of D3 feature**Box Plots of D3 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	327662.000000	count	318043.000000	9619.000000
mean	28.343348	mean	28.845357	11.744880
std	62.384721	std	62.830058	42.005468
min	0.000000	min	0.000000	0.000000
25%	1.000000	25%	1.000000	0.000000
50%	8.000000	50%	8.000000	1.000000
75%	27.000000	75%	28.000000	5.000000
max	819.000000	max	819.000000	582.000000
Name: D3, dtype: float64		Name: D3, dtype: float64		Name: D3, dtype: float64

The unique values in D3 feature is: 650

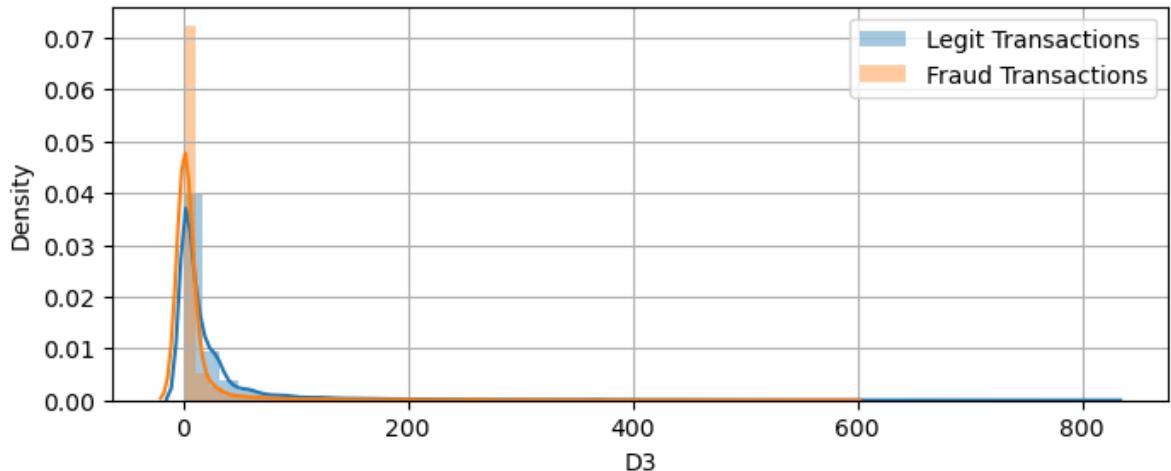
D3 feature has 262878 null values, which holds 0.44514850814508755 %

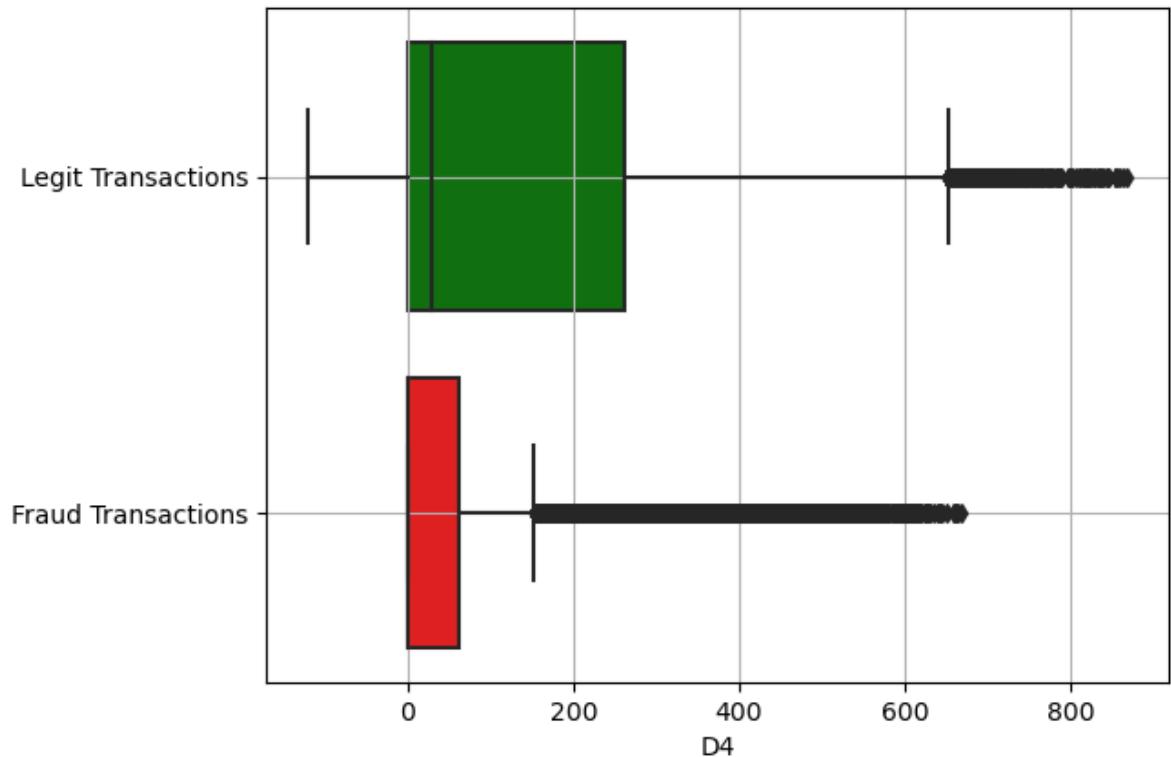
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	1.0	0.0
30	2.0	0.0
35	3.0	0.0
40	5.0	0.0
45	7.0	0.0
50	8.0	1.0
55	11.0	1.0
60	14.0	1.0
65	17.0	2.0
70	21.0	3.0
75	28.0	5.0
80	32.0	7.0
85	43.0	13.0
90	67.0	22.0
95	131.0	54.0
100	819.0	582.0

InterQuartileRange of legit transactions: 27.0

InterQuartileRange of fraud transactions: 5.0

D3 of Legit Transaction and Fraud Transaction

Analysis of D4 feature**Box Plots of D4 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	421618.000000	count	407051.000000	count
mean	140.002441	mean	142.432339	mean
std	191.096774	std	192.213375	std
min	-122.000000	min	-122.000000	min
25%	0.000000	25%	0.000000	25%
50%	26.000000	50%	28.000000	50%
75%	253.000000	75%	261.000000	75%
max	869.000000	max	869.000000	max
	Name: D4, dtype: float64		Name: D4, dtype: float64	
	Name: D4, dtype: float64			

The unique values in D4 feature is: 809

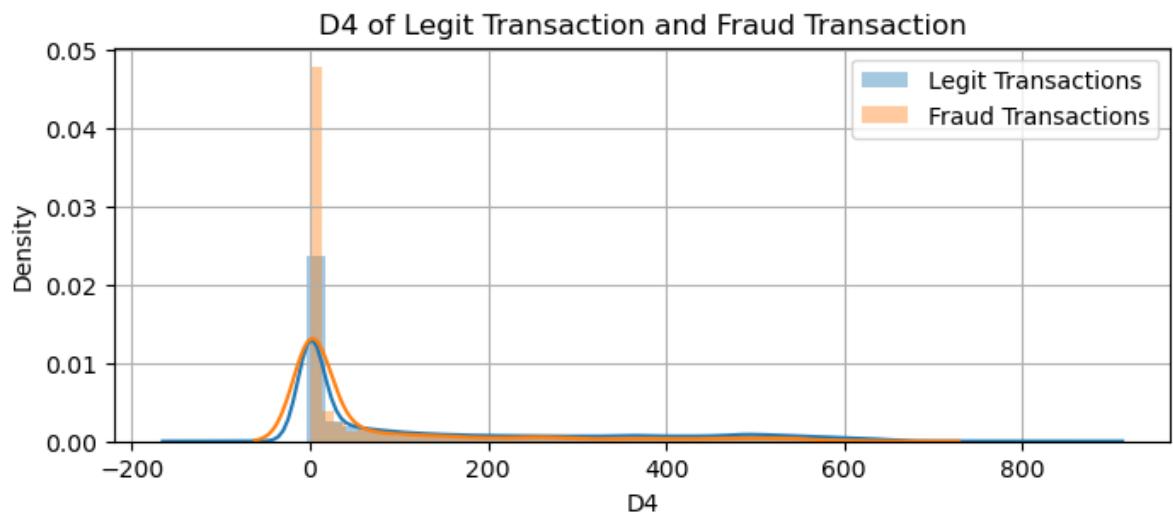
D4 feature has 168922 null values, which holds 0.2860466691502693 %

Percentiles

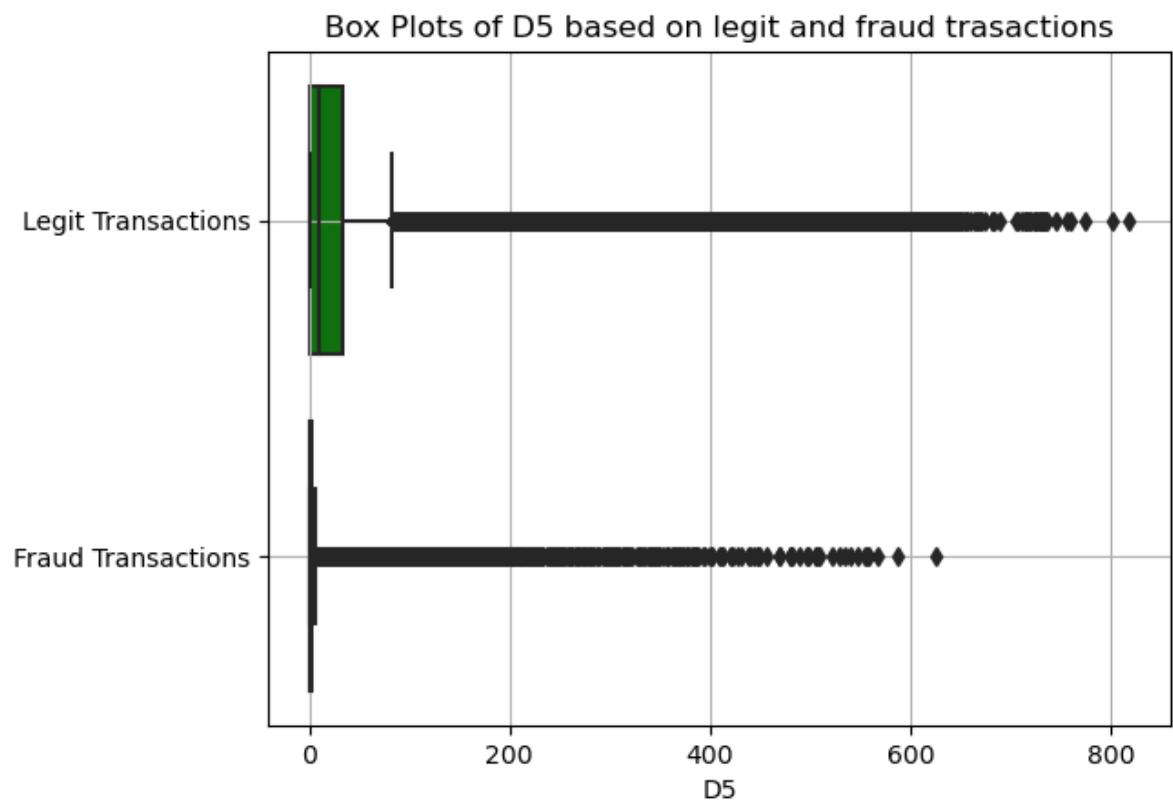
Percentile	Legit Transactions	Fraud Transactions
0	-122.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	1.0	0.0
45	11.0	0.0
50	28.0	1.0
55	52.0	3.0
60	85.0	7.0
65	128.0	16.0
70	188.0	31.0
75	261.0	61.0
80	338.0	112.0
85	412.0	188.0
90	484.0	289.4
95	546.0	453.0
100	869.0	667.0

InterQuartileRange of legit transactions: 261.0

InterQuartileRange of fraud transactions: 61.0



Analysis of D5 feature



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	280699.000000	count	269819.000000	count
mean	42.335965	mean	43.491155	mean
std	89.000144	std	89.985719	std
min	0.000000	min	0.000000	min
25%	1.000000	25%	1.000000	25%
50%	0.000000	50%	10.000000	50%
75%	32.000000	75%	34.000000	75%
max	819.000000	max	819.000000	max
	625.000000			
Name: D5, dtype: float64		Name: D5, dtype: float64		
Name: D5, dtype: float64				

The unique values in D5 feature is: 689

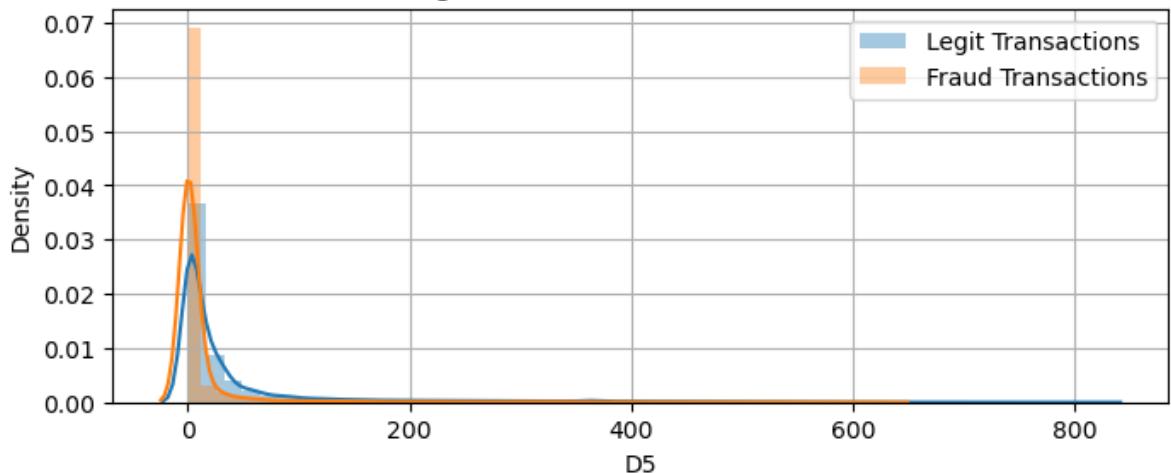
D5 feature has 309841 null values, which holds 0.524674027161581 %

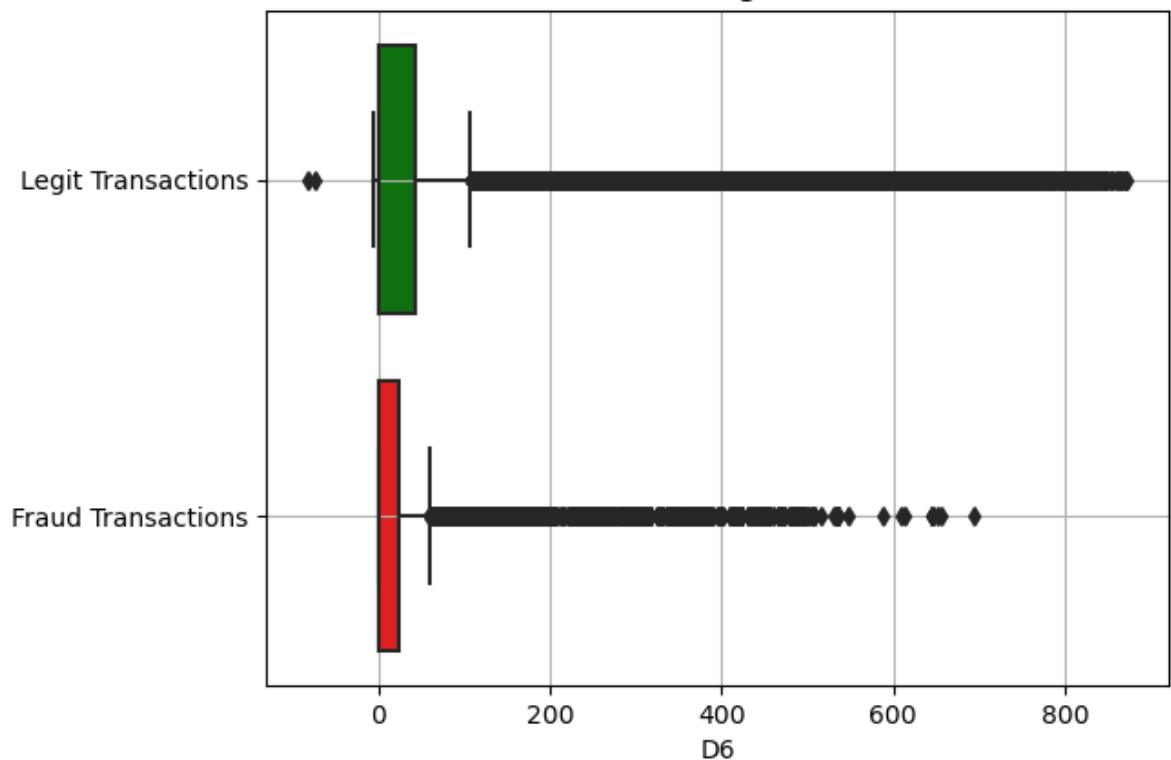
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	1.0	0.0
30	2.0	0.0
35	4.0	0.0
40	6.0	0.0
45	8.0	0.0
50	10.0	0.0
55	14.0	0.0
60	16.0	0.0
65	22.0	1.0
70	28.0	1.0
75	34.0	3.0
80	47.0	5.0
85	70.0	10.0
90	121.0	23.0
95	251.0	69.0
100	819.0	625.0

InterQuartileRange of legit transactions: 33.0

InterQuartileRange of fraud transactions: 3.0

D5 of Legit Transaction and Fraud Transaction

Analysis of D6 feature**Box Plots of D6 based on legit and fraud transactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	73187.000000	count	65469.000000	
mean	69.805717	mean	72.629046	
std	143.669253	std	147.600963	
min	-83.000000	min	-83.000000	
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	40.000000	75%	43.000000	
max	873.000000	max	873.000000	
	Name: D6, dtype: float64		Name: D6, dtype: float64	
	Name: D6, dtype: float64			

The unique values in D6 feature is: 830

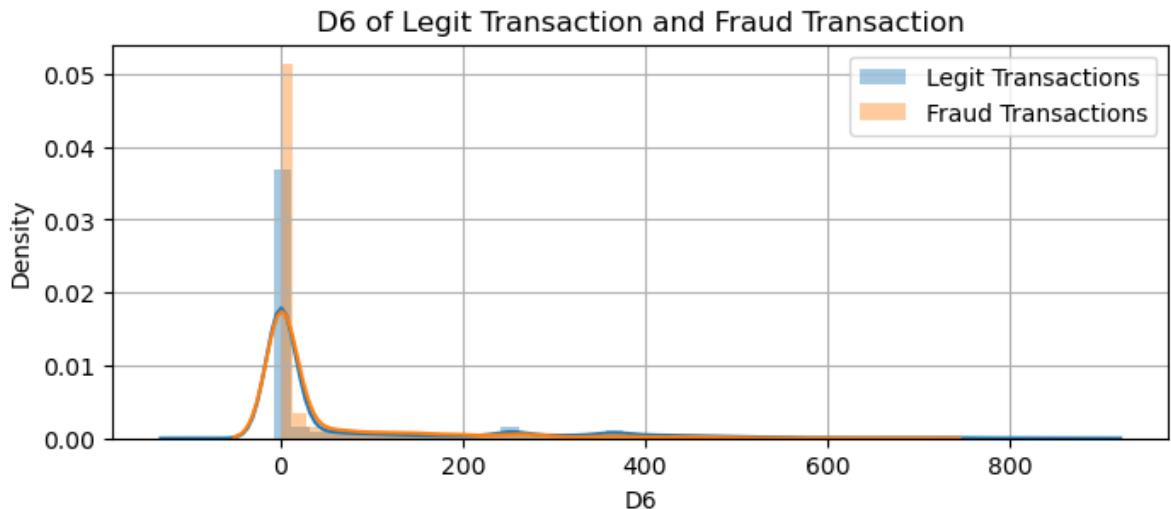
D6 feature has 517353 null values, which holds 0.8760676668811597 %

Percentiles

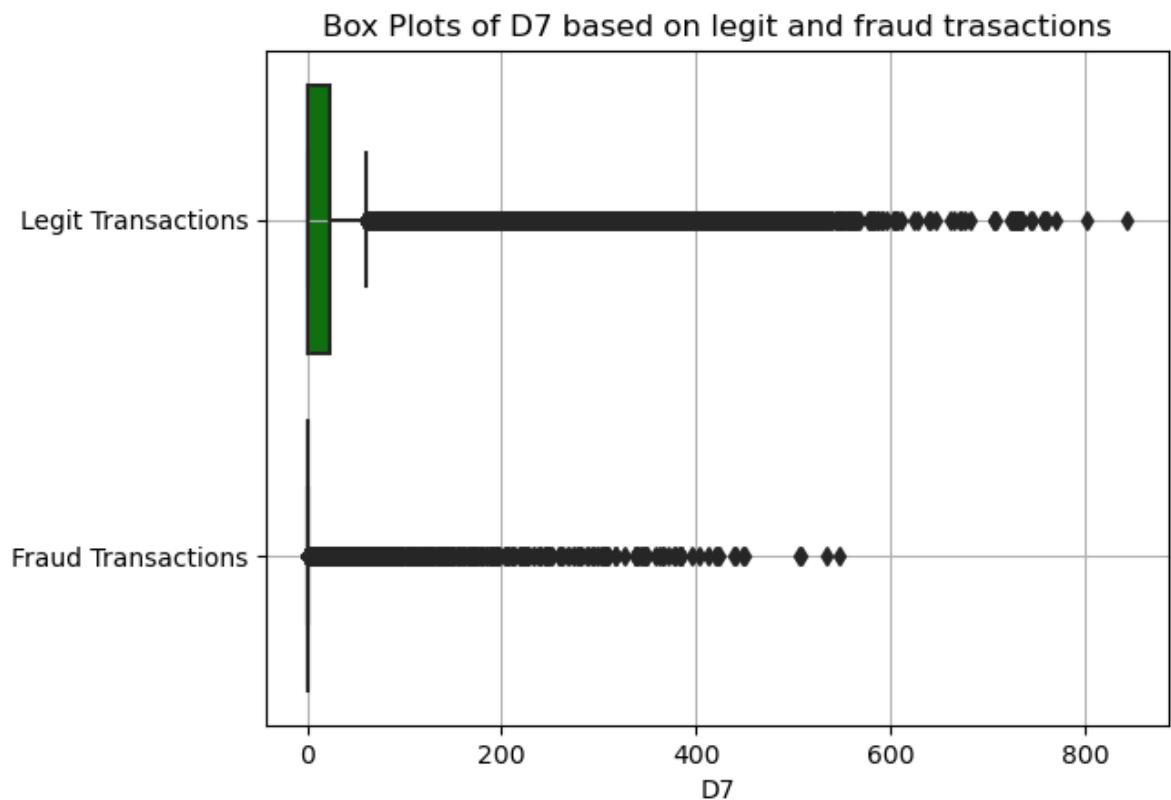
Percentile	Legit Transactions	Fraud Transactions
0	-83.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	1.0
65	1.0	4.0
70	10.0	10.0
75	43.0	24.0
80	115.0	57.0
85	249.0	112.0
90	328.0	186.3
95	413.0	283.0
100	873.0	694.0

InterQuartileRange of legit transactions: 43.0

InterQuartileRange of fraud transactions: 24.0



Analysis of D7 feature



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	38917.000000	count	33127.000000	
count	5790.000000			
mean	41.638950	mean	46.943007	
mean	11.292228			
std	99.743264	std	105.435756	
std	46.743057			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	17.000000	75%	24.000000	
75%	0.000000			
max	843.000000	max	843.000000	
max	547.000000			
Name: D7, dtype: float64		Name: D7, dtype: float64		
Name: D7, dtype: float64				

The unique values in D7 feature is: 598

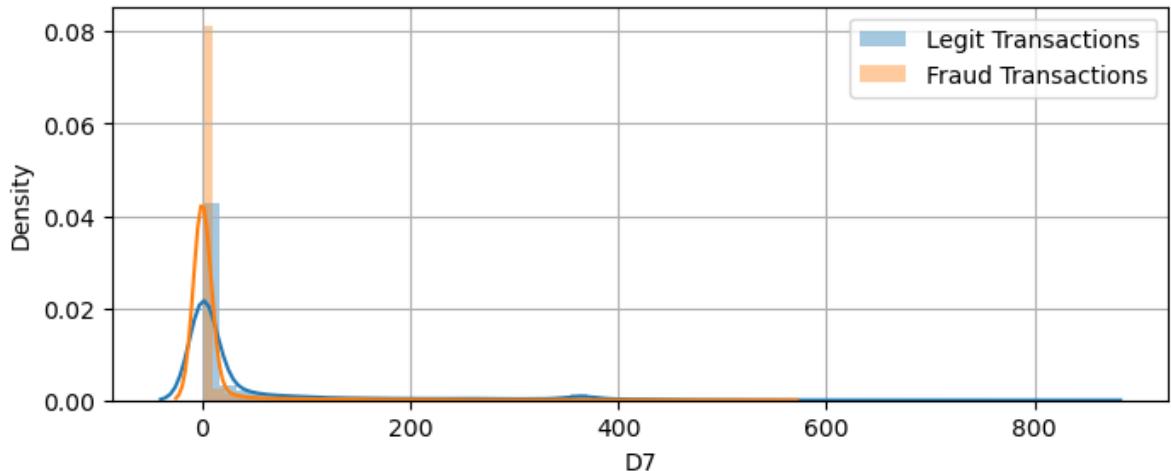
D7 feature has 551623 null values, which holds 0.9340992989467267 %

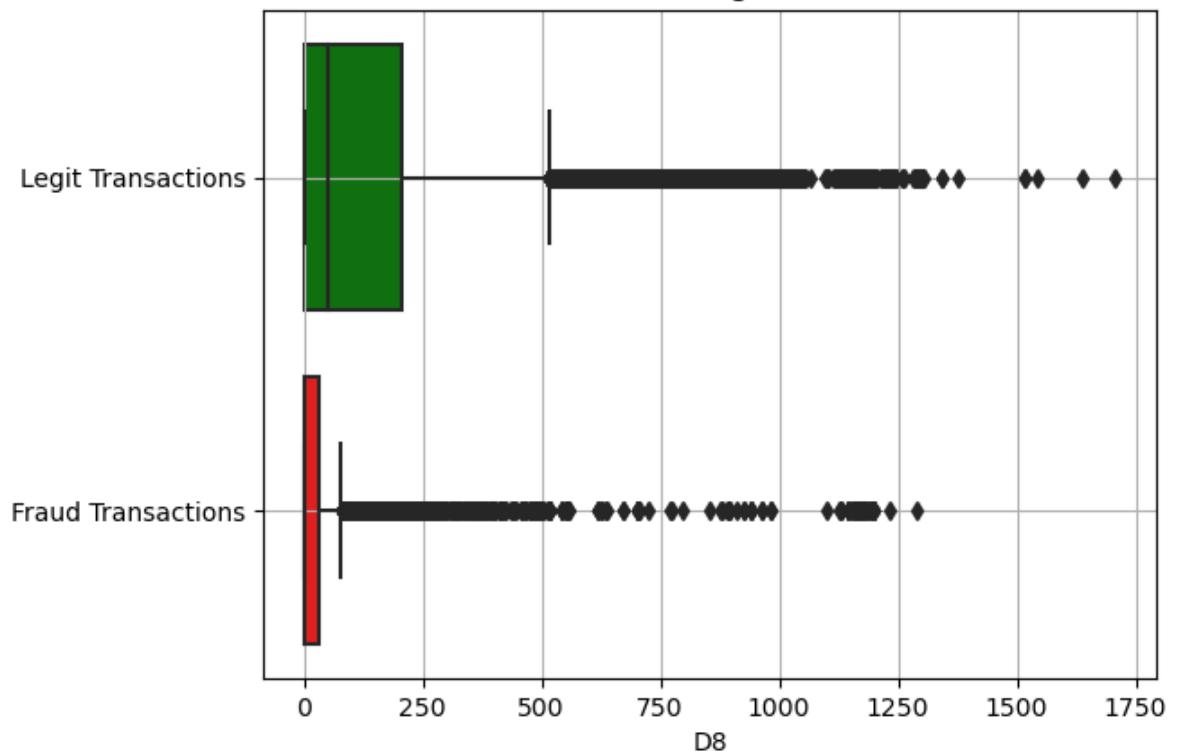
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	2.0	0.0
60	4.0	0.0
65	7.0	0.0
70	13.0	0.0
75	24.0	0.0
80	46.0	1.0
85	90.0	4.0
90	189.0	14.0
95	348.0	63.55
100	843.0	547.0

InterQuartileRange of legit transactions: 24.0

InterQuartileRange of fraud transactions: 0.0

D7 of Legit Transaction and Fraud Transaction

Analysis of D8 feature**Box Plots of D8 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	74926.000000		67099.000000	
mean	146.058108		mean	157.343727
std	231.663840		std	237.987816
min	0.000000		min	0.000000
25%	0.958333		25%	1.750000
50%	0.750000		50%	50.583332
75%	187.958328		75%	207.541672
max	1707.791626		max	1707.791626
	1290.583374			
Name: D8, dtype: float64		Name: D8, dtype: float64		
Name: D8, dtype: float64				

The unique values in D8 feature is: 12354

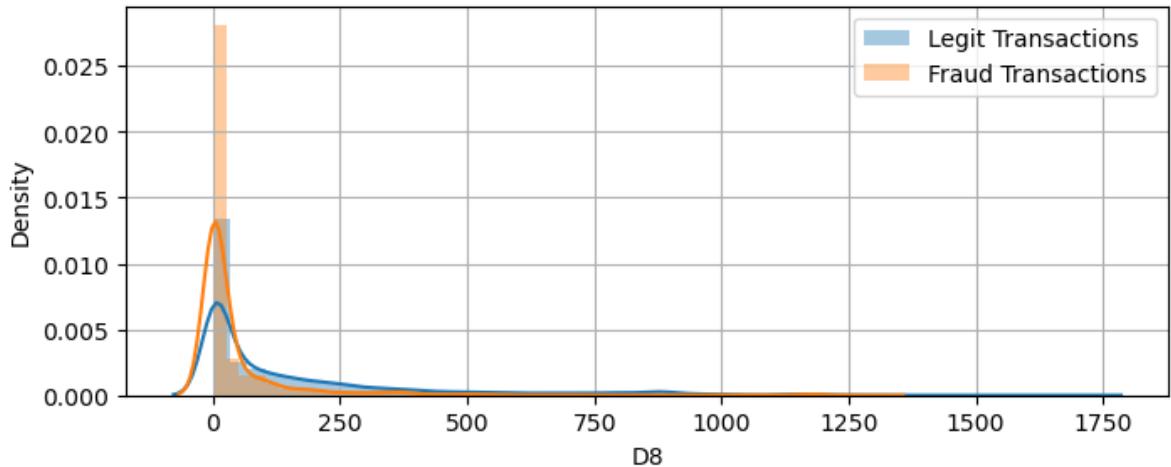
D8 feature has 515614 null values, which holds 0.8731229044603245 %

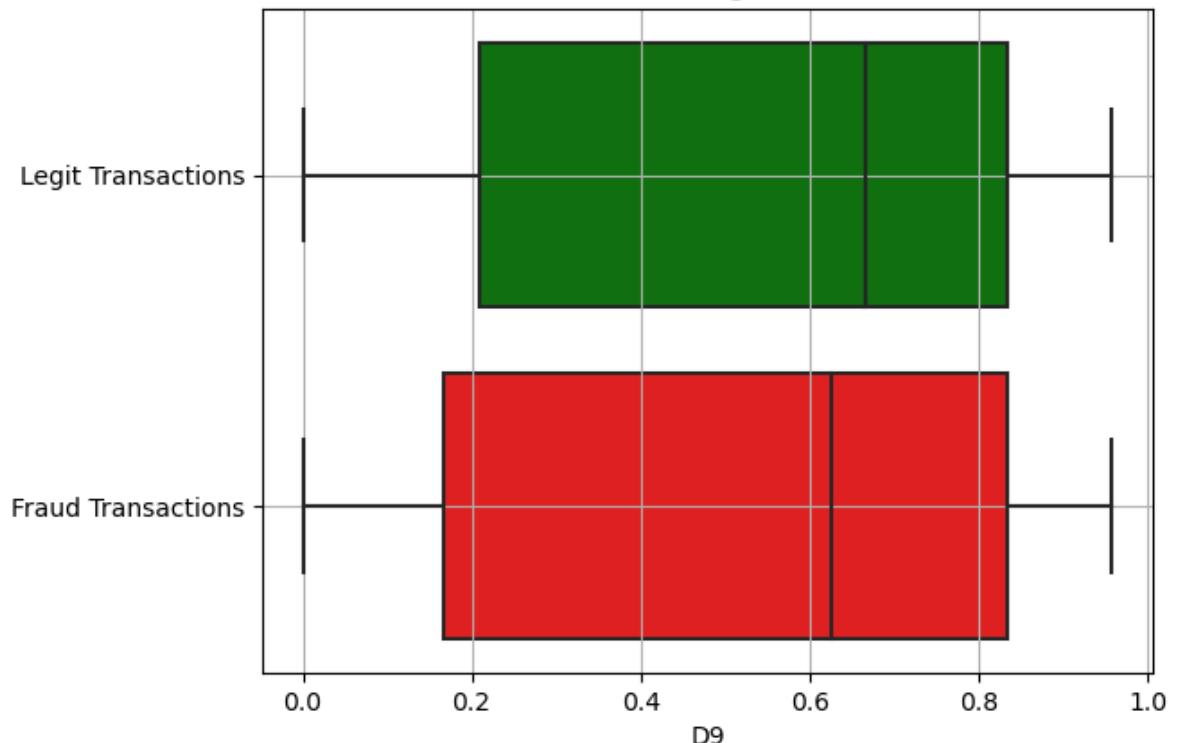
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.167	0.083
10	0.625	0.167
15	0.75	0.292
20	0.875	0.625
25	1.75	0.75
30	6.0	0.833
35	12.708	0.875
40	21.625	0.958
45	31.667	1.167
50	50.583	2.667
55	71.542	4.125
60	96.917	7.125
65	127.542	11.167
70	162.875	19.667
75	207.542	32.375
80	260.583	52.167
85	340.804	87.917
90	478.542	131.833
95	768.667	268.017
100	1707.792	1290.583

InterQuartileRange of legit transactions: 205.792

InterQuartileRange of fraud transactions: 31.625

D8 of Legit Transaction and Fraud Transaction

Analysis of D9 feature**Box Plots of D9 based on legit and fraud transactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count		count		count
count	74926.000000		67099.000000	
mean	0.561057		mean	0.565846
std	0.316880		std	0.314576
min	0.000000		min	0.000000
25%	0.208333		25%	0.208333
50%	0.166666		50%	0.666666
75%	0.625000		75%	0.833333
max	0.833333		max	0.958333
Name: D9, dtype: float64		Name: D9, dtype: float64		
Name: D9, dtype: float64				

The unique values in D9 feature is: 25

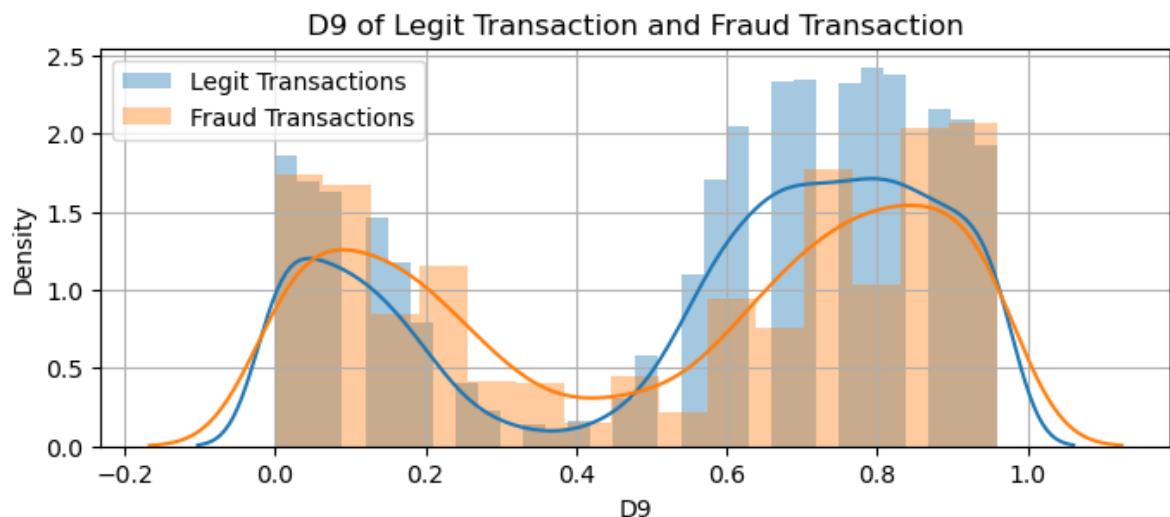
D9 feature has 515614 null values, which holds 0.8731229044603245 %

Percentiles

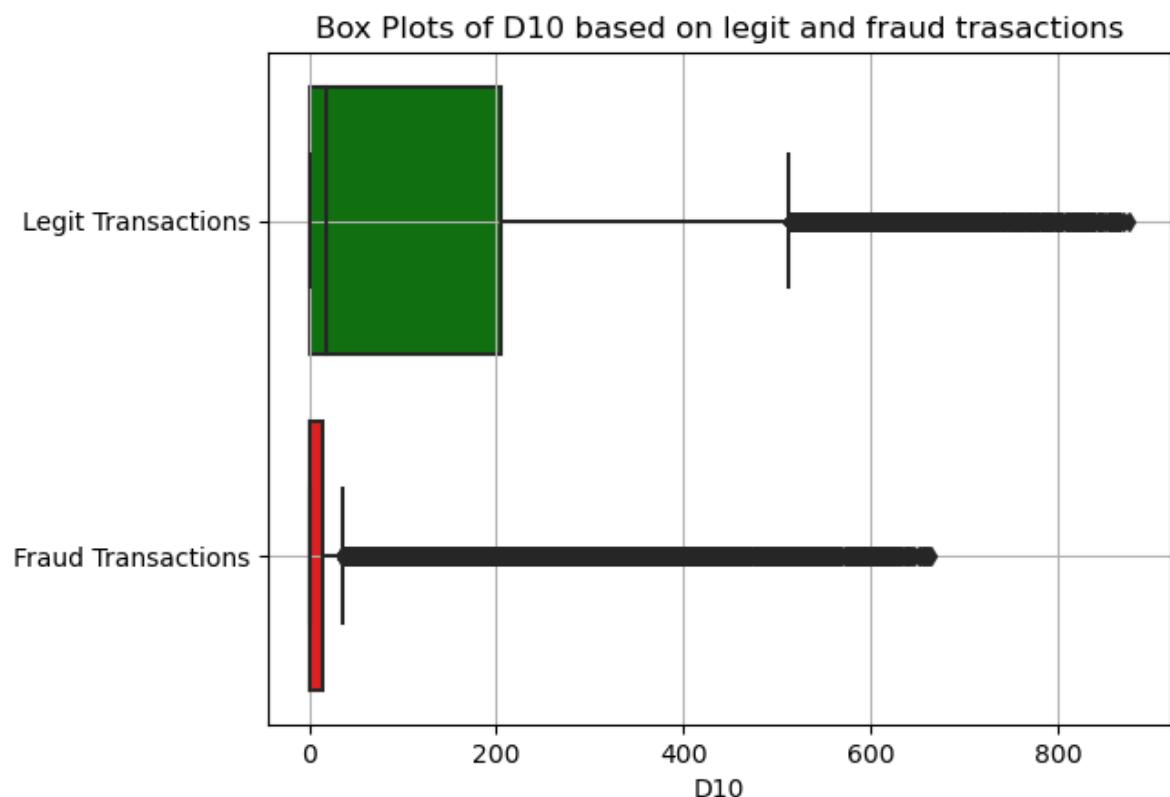
Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.042	0.042
15	0.083	0.083
20	0.167	0.125
25	0.208	0.167
30	0.5	0.208
35	0.583	0.292
40	0.583	0.417
45	0.625	0.583
50	0.667	0.625
55	0.708	0.667
60	0.708	0.708
65	0.75	0.75
70	0.792	0.792
75	0.833	0.833
80	0.833	0.833
85	0.875	0.875
90	0.917	0.917
95	0.958	0.958
100	0.958	0.958

InterQuartileRange of legit transactions: 0.625

InterQuartileRange of fraud transactions: 0.6659999999999999



Analysis of D10 feature



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	514518.000000	count	497741.000000	
mean	123.982137	mean	126.396130	
std	52.363653	std	183.710646	
min	0.000000	min	0.000000	
25%	0.000000	25%	0.000000	
50%	15.000000	50%	18.000000	
75%	197.000000	75%	205.000000	
max	876.000000	max	876.000000	
Name: D10, dtype: float64		Name: D10, dtype: float64		
Name: D10, dtype: float64				

The unique values in D10 feature is: 819

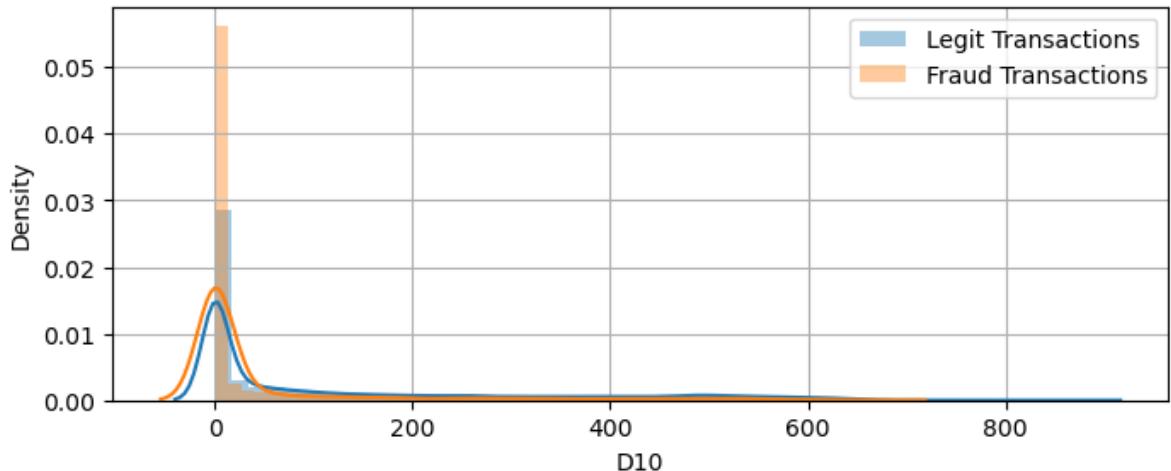
D10 feature has 76022 null values, which holds 0.1287330240119213 %

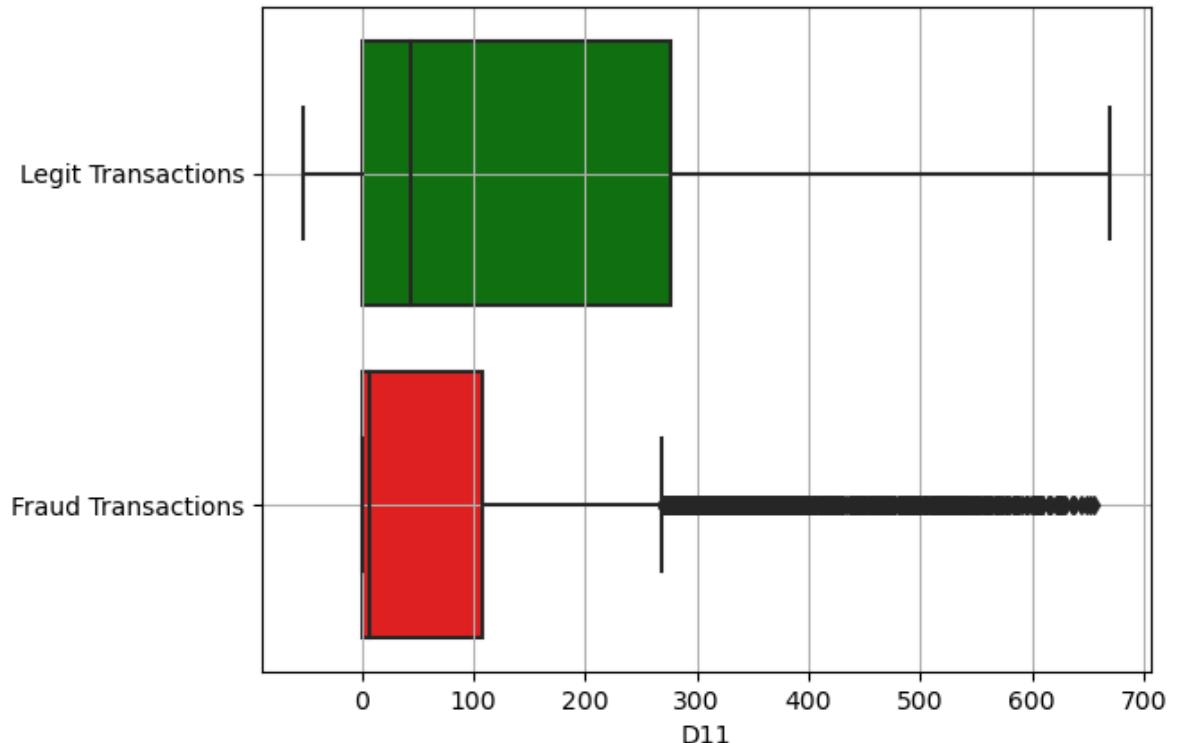
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	5.0	0.0
50	18.0	0.0
55	35.0	0.0
60	61.0	0.0
65	94.0	1.0
70	142.0	4.0
75	205.0	14.0
80	279.0	38.0
85	370.0	96.0
90	465.0	207.0
95	537.0	398.2
100	876.0	664.0

InterQuartileRange of legit transactions: 205.0

InterQuartileRange of fraud transactions: 14.0

D10 of Legit Transaction and Fraud Transaction

Analysis of D11 feature**Box Plots of D11 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	311253.000000	count	305147.000000	
mean	146.621465	mean	147.808201	
std	186.042622	std	186.555365	
min	-53.000000	min	-53.000000	
25%	0.000000	25%	0.000000	
50%	43.000000	50%	44.000000	
75%	274.000000	75%	277.000000	
max	107.750000	max	670.000000	
		Name: D11, dtype: float64	Name: D11, dtype: float64	
	Name: D11, dtype: float64			

The unique values in D11 feature is: 677

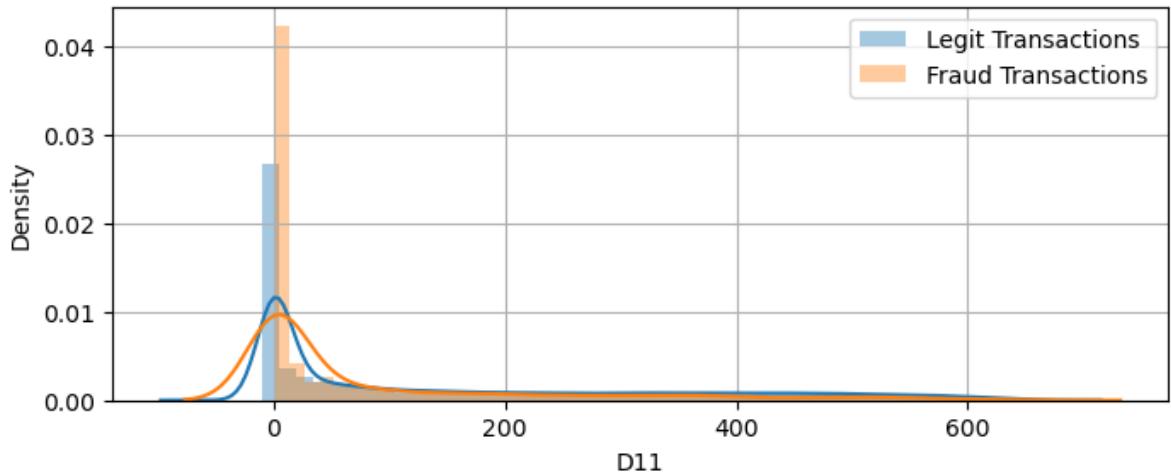
D11 feature has 279287 null values, which holds 0.47293494090154775 %

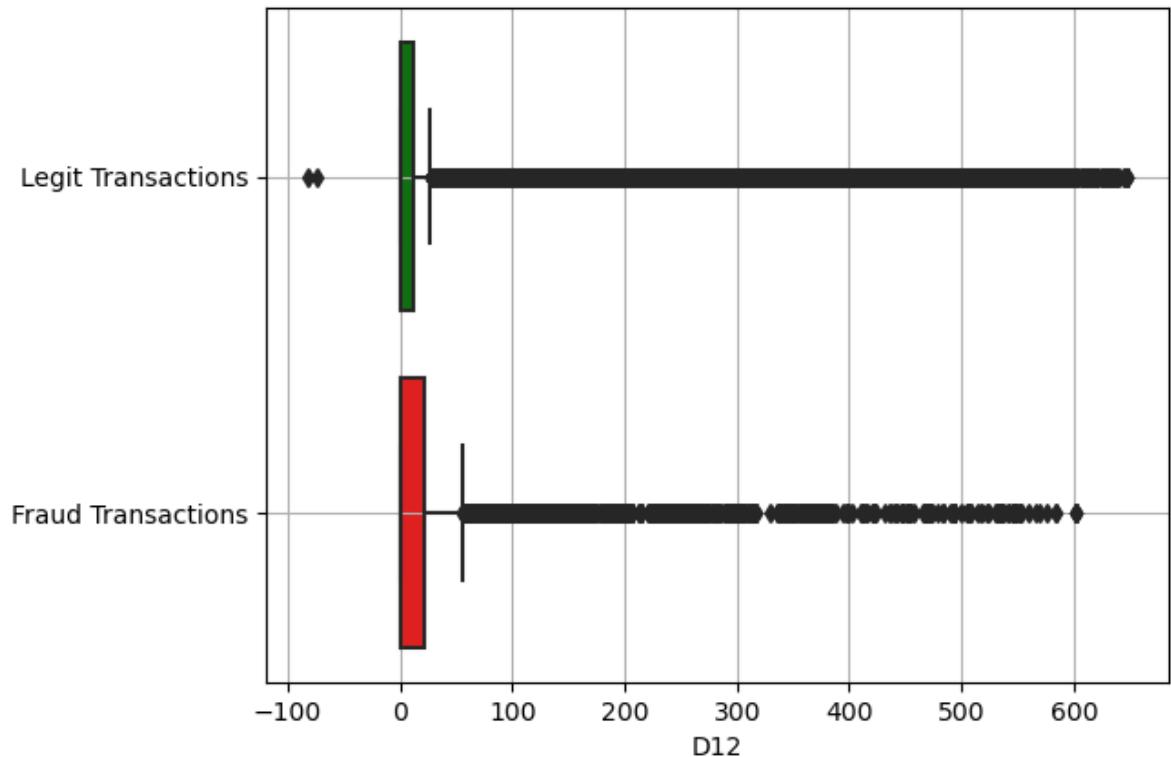
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	-53.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	8.0	1.0
45	24.0	3.0
50	44.0	7.0
55	73.0	12.0
60	112.0	24.0
65	158.0	44.0
70	215.0	68.0
75	277.0	107.75
80	339.0	171.0
85	401.0	241.25
90	464.0	334.0
95	533.0	449.0
100	670.0	655.0

InterQuartileRange of legit transactions: 277.0

InterQuartileRange of fraud transactions: 107.75

D11 of Legit Transaction and Fraud Transaction

Analysis of D12 feature**Box Plots of D12 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	64717.000000	count	57119.000000	
mean	54.037533	mean	55.345787	
std	124.274558	std	127.166079	
min	-83.000000	min	-83.000000	
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	13.000000	75%	11.000000	
max	648.000000	max	648.000000	
Name: D12, dtype: float64		Name: D12, dtype: float64		
Name: D12, dtype: float64				

The unique values in D12 feature is: 636

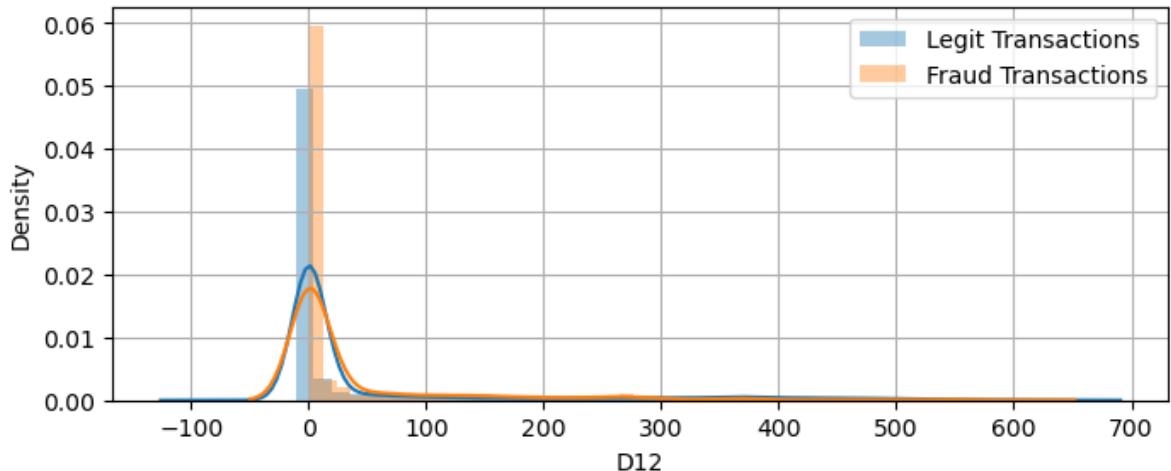
D12 feature has 525823 null values, which holds 0.8904104717715988 %

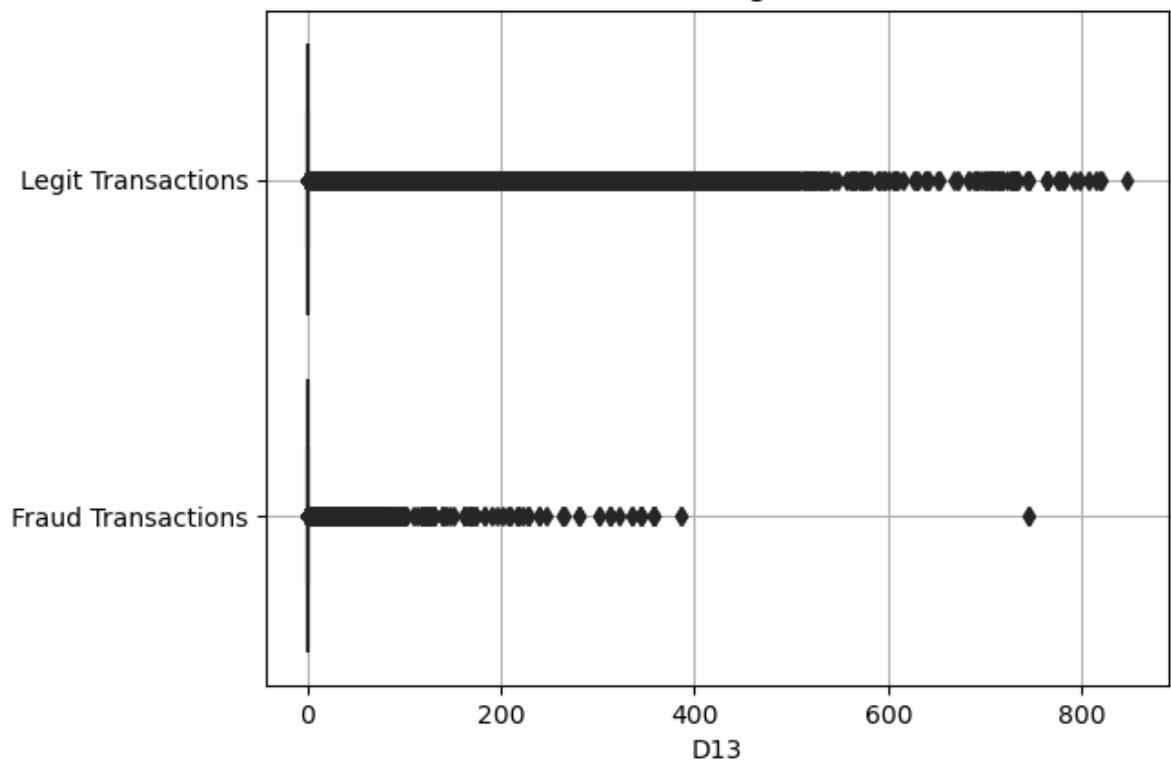
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	-83.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	1.0
65	0.0	3.0
70	2.0	9.0
75	11.0	22.0
80	44.0	55.0
85	119.0	106.45
90	273.0	174.0
95	390.0	269.0
100	648.0	602.0

InterQuartileRange of legit transactions: 11.0

InterQuartileRange of fraud transactions: 22.0

D12 of Legit Transaction and Fraud Transaction

Analysis of D13 feature**Box Plots of D13 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	61952.000000	count	55115.000000	
count	6837.000000			
mean	17.901295	mean	19.316574	
mean	6.492321			
std	67.614425	std	70.703174	
std	31.330429			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	847.000000	max	847.000000	
max	745.000000			
Name: D13, dtype: float64		Name: D13, dtype: float64		
Name: D13, dtype: float64				

The unique values in D13 feature is: 578

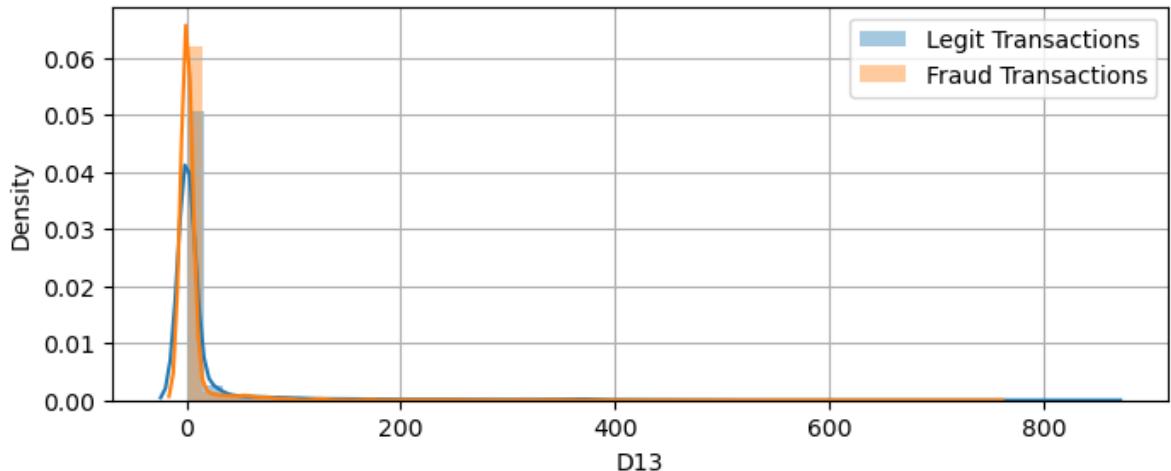
D13 feature has 528588 null values, which holds 0.8950926270870728 %

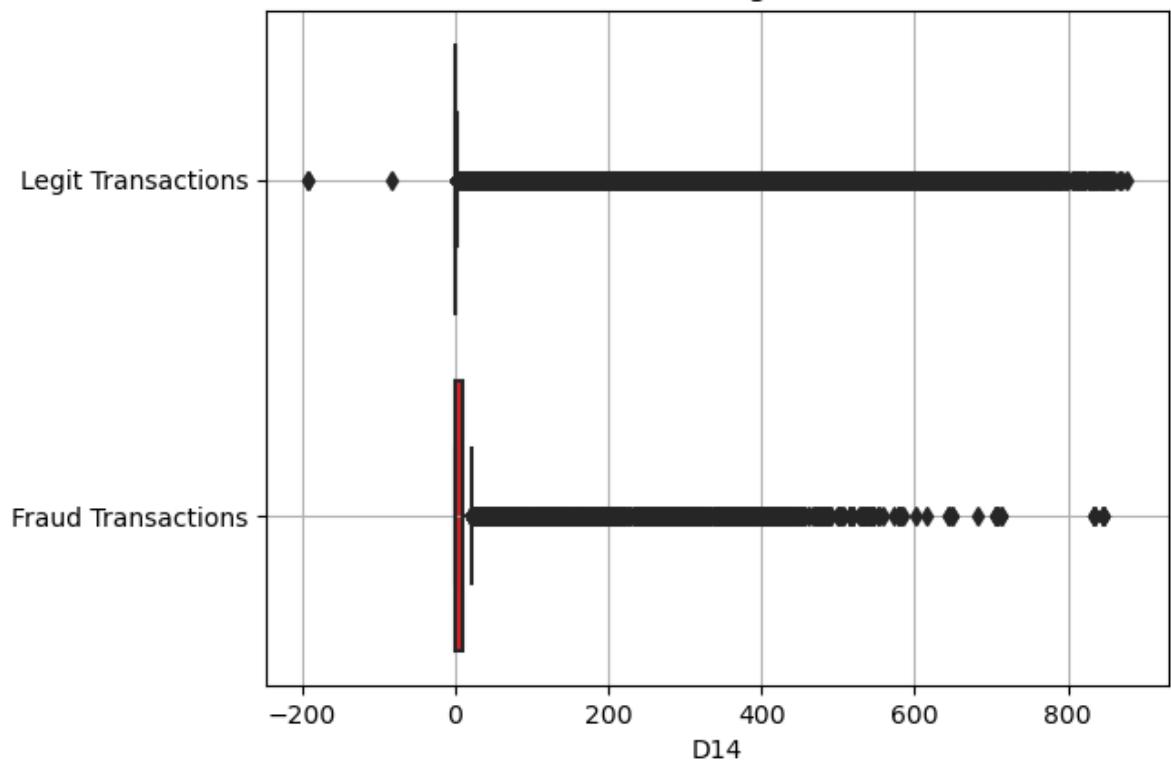
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	0.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	0.0
80	1.0	1.0
85	13.0	2.0
90	31.0	7.0
95	123.0	36.0
100	847.0	745.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 0.0

D13 of Legit Transaction and Fraud Transaction

Analysis of D14 feature**Box Plots of D14 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	62187.000000	count	54974.000000	
mean	57.724444	mean	58.152163	
std	136.312450	std	137.532930	
min	-193.000000	min	-193.000000	
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	2.000000	75%	1.000000	
max	878.000000	max	878.000000	
	Name: D14, dtype: float64		Name: D14, dtype: float64	
	Name: D14, dtype: float64			

The unique values in D14 feature is: 803

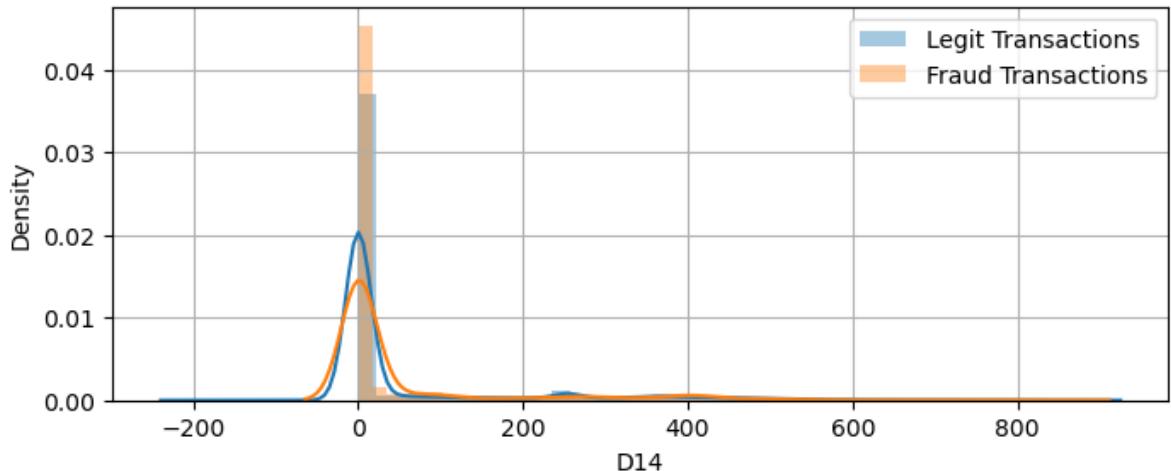
D14 feature has 528353 null values, which holds 0.8946946862193924 %

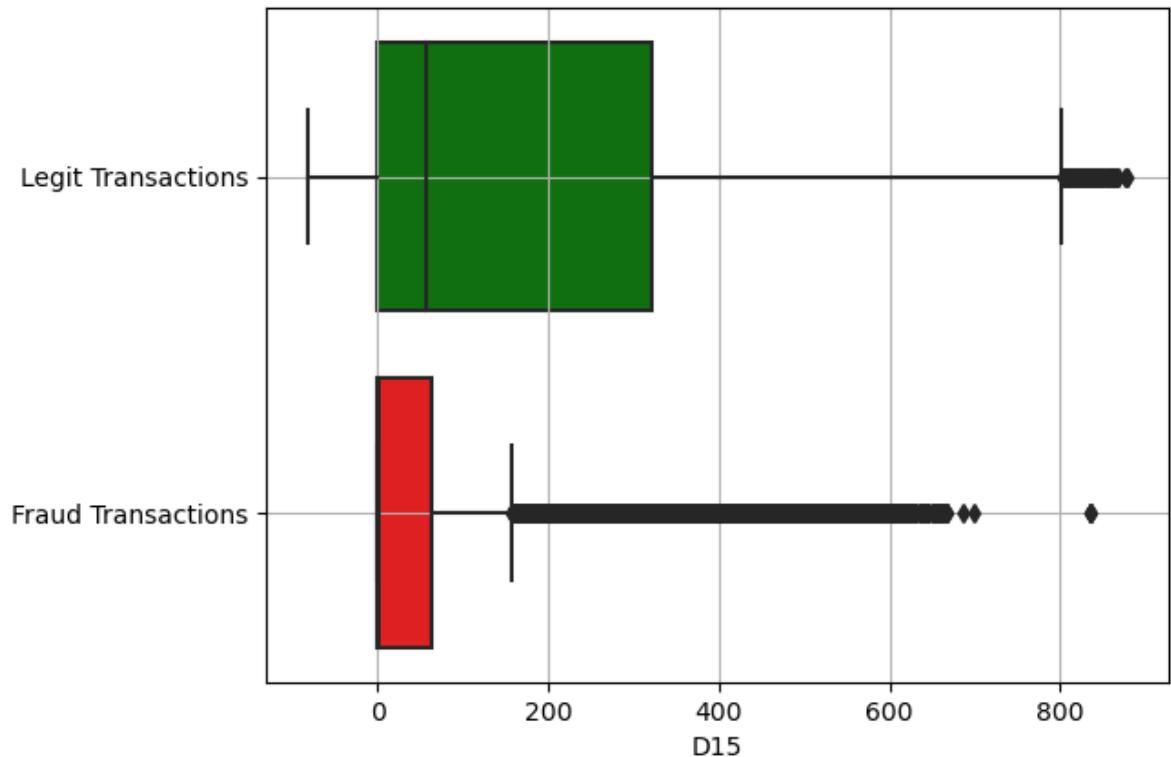
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	-193.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	1.0
70	0.0	3.0
75	1.0	9.0
80	31.0	42.0
85	149.0	107.6
90	261.0	269.0
95	397.0	396.0
100	878.0	847.0

InterQuartileRange of legit transactions: 1.0

InterQuartileRange of fraud transactions: 9.0

D14 of Legit Transaction and Fraud Transaction

Analysis of D15 feature**Box Plots of D15 based on legit and fraud trasactions**

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	501427.000000	count	485040.000000	
mean	163.744579	mean	166.633142	
std	78.245805	std	203.595694	
min	-83.000000	min	-83.000000	
25%	0.000000	25%	0.000000	
50%	52.000000	50%	56.000000	
75%	1.000000	75%	321.000000	
max	314.000000	max	879.000000	
	63.000000			
	835.000000			
	Name: D15, dtype: float64		Name: D15, dtype: float64	
	Name: D15, dtype: float64			

The unique values in D15 feature is: 860

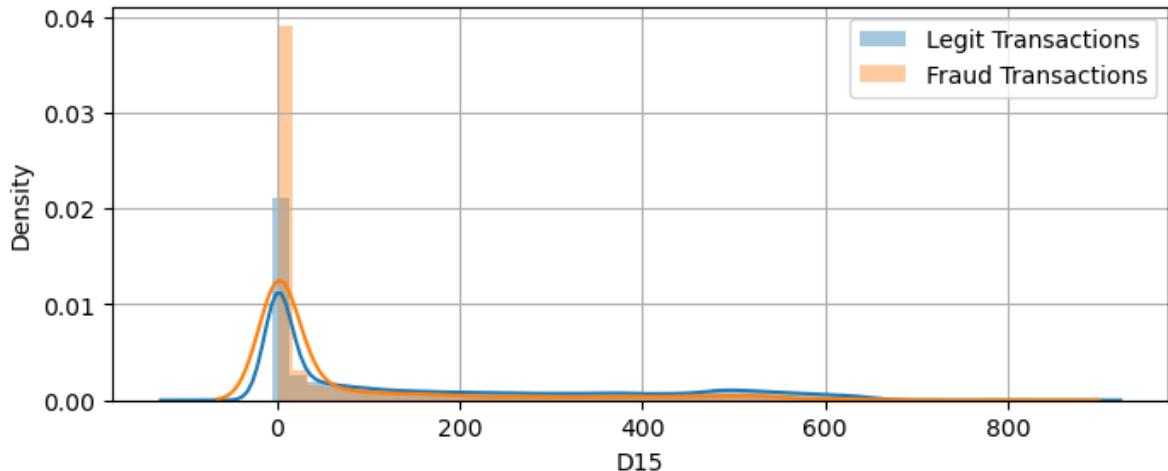
D15 feature has 89113 null values, which holds 0.1509008703898127 %

Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	-83.0	0.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	1.0	0.0
40	12.0	0.0
45	30.0	0.0
50	56.0	1.0
55	90.0	2.0
60	134.0	7.0
65	188.0	16.0
70	251.0	32.0
75	321.0	63.0
80	389.0	120.0
85	460.0	199.1
90	511.0	350.4
95	568.0	478.0
100	879.0	835.0

InterQuartileRange of legit transactions: 321.0

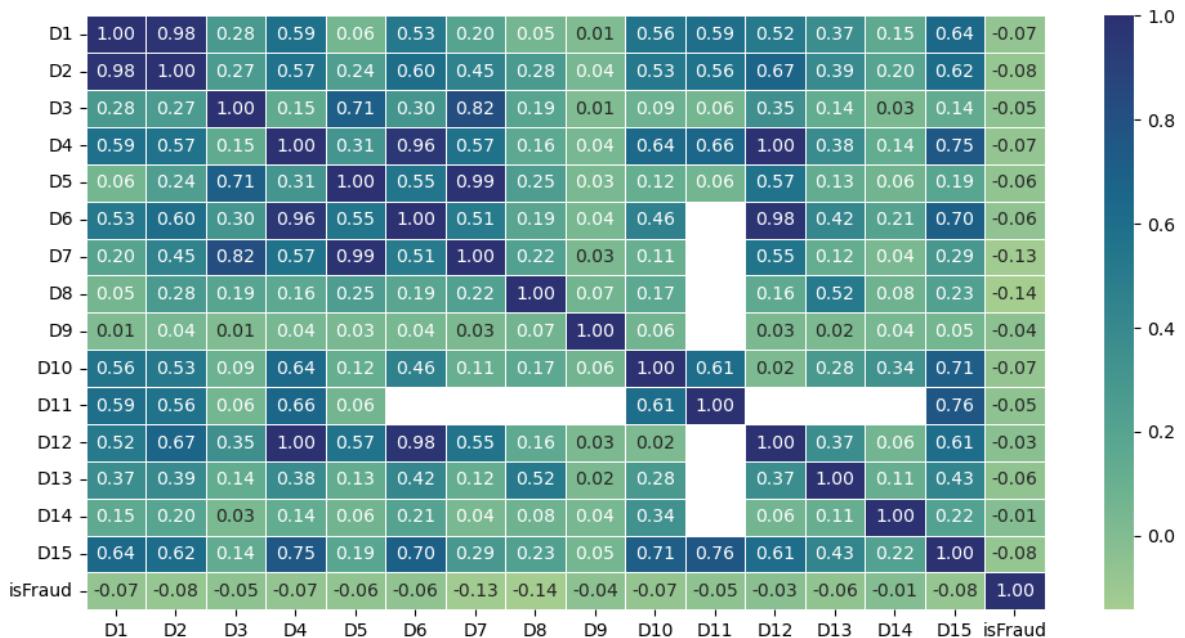
InterQuartileRange of fraud transactions: 63.0

D15 of Legit Transaction and Fraud Transaction

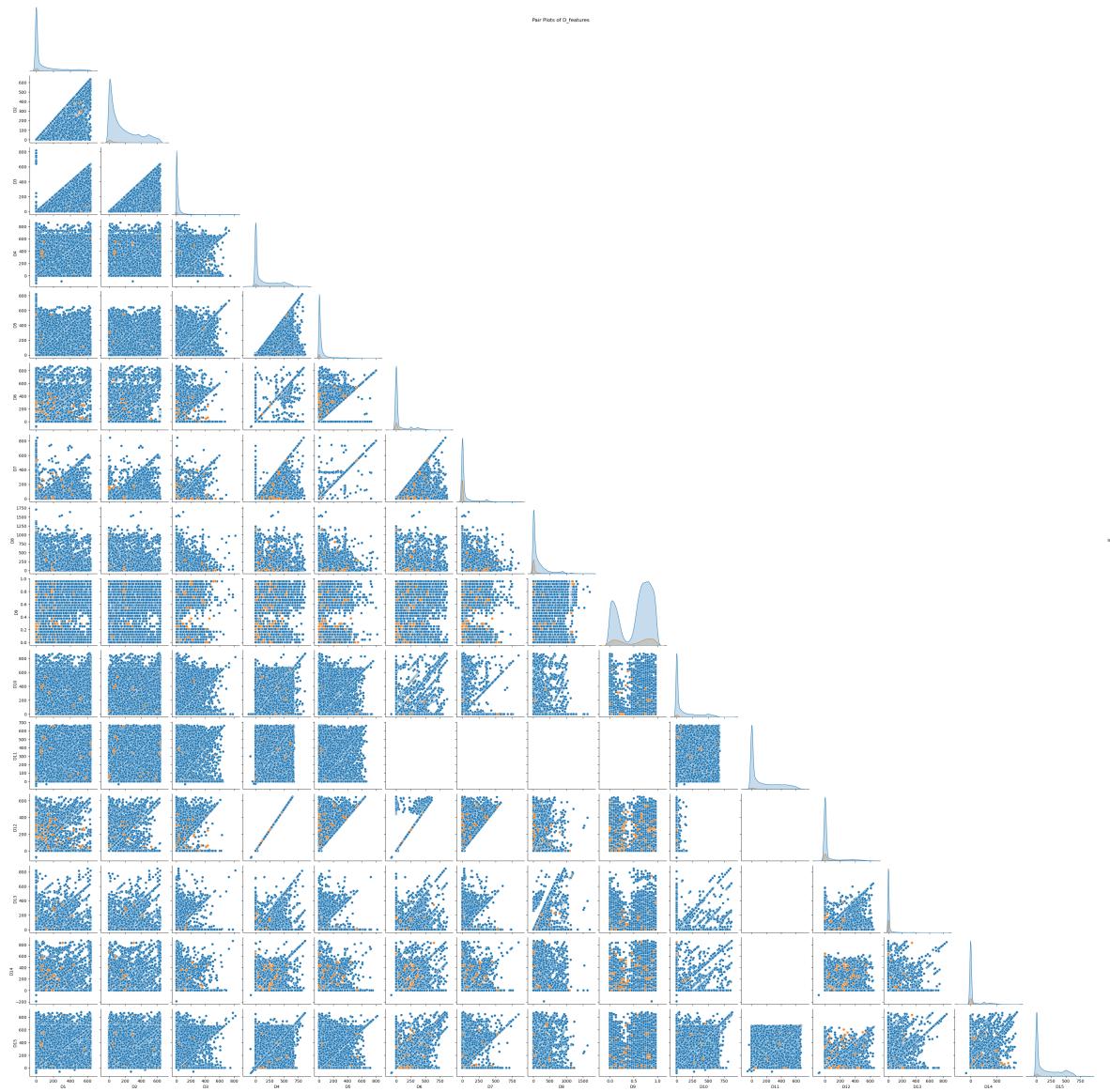
```
*****
*****
```

In [77]: `# to get collinearity of D features`

```
plt.figure(figsize=(12,6))
sns.heatmap(df_train[d_features + ['isFraud']].corr(), annot = True, fmt=".2f", cmap='viridis')
plt.show()
```



In [78]: `sns.pairplot(vars = d_features, hue='isFraud', data = df_train, corner=True)`
`plt.suptitle("\nPair Plots of D_features\n")`
`plt.show()`



Observation:

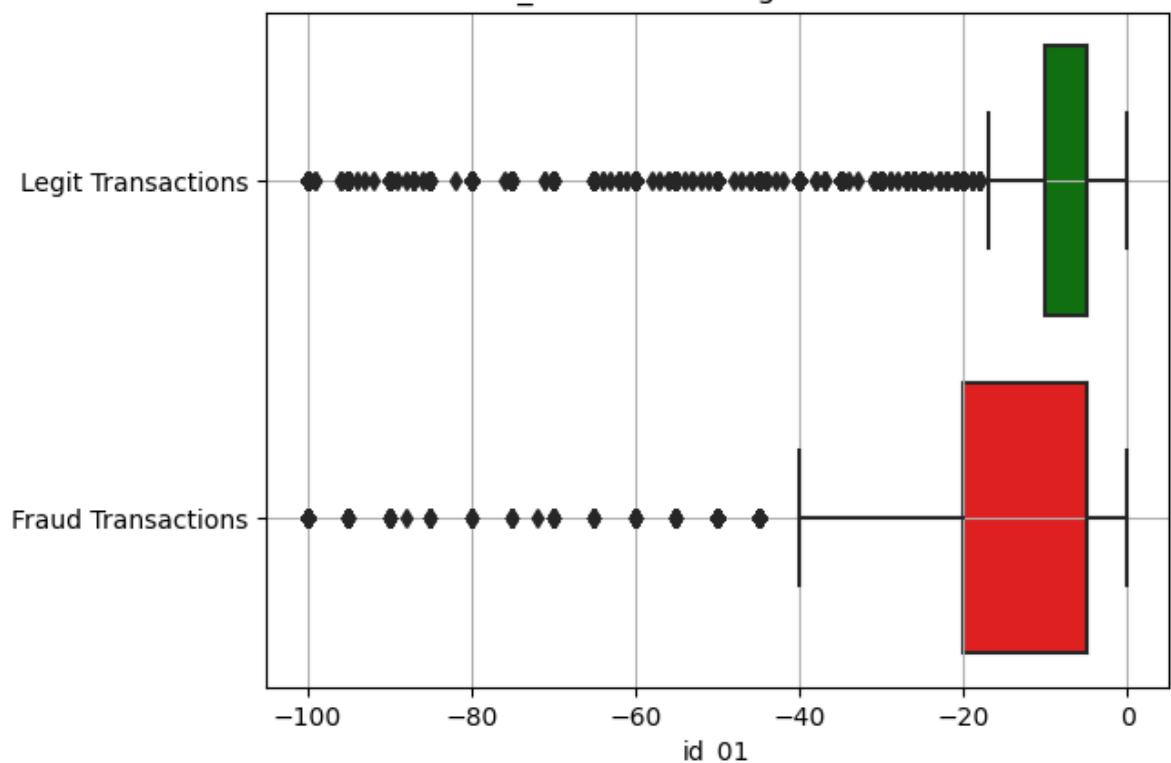
1. D_features relate to the timedelta, such as the number of days since the last transaction, etc.
2. It has been noted that the gap between legitimate transactions appears to be larger as a result.
3. In contrast, fraudulent cases occur more frequently.
4. For this group of characteristics, there are a few correlated features as well, which we will handle later.

1.2.6 Feature: 'ID features'

```
In [79]: id_features = ['id_01', 'id_02', 'id_03', 'id_04', 'id_05', 'id_06', 'id_07', 'id_08']
for fea in id_features:
    print(f'Analysis of {fea} feature')
    analysis_num_fea(df_train, fea, 'isFraud')
    print('\n')
    print('*'*120)
    print('\n')
```

Analysis of id_01 feature

Box Plots of id_01 based on legit and fraud trasactions



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	144233.000000	count	132915.000000	
mean	-10.170502	mean	-9.667667	
std	14.347949	std	13.592128	
min	-100.000000	min	-100.000000	
25%	-10.000000	25%	-10.000000	
50%	-5.000000	50%	-5.000000	
75%	-5.000000	75%	-5.000000	
max	0.000000	max	0.000000	
	Name: id_01, dtype: float64		Name: id_01, dtype: float64	
	Name: id_01, dtype: float64			

The unique values in id_01 feature is: 78

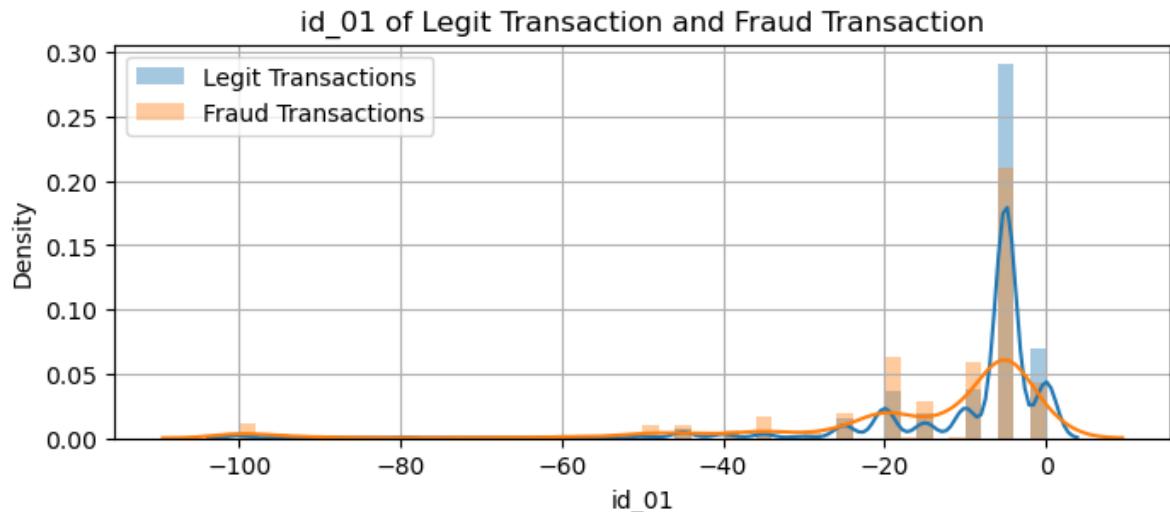
id_01 feature has 446307 null values, which holds 0.7557608290716971 %

Percentiles

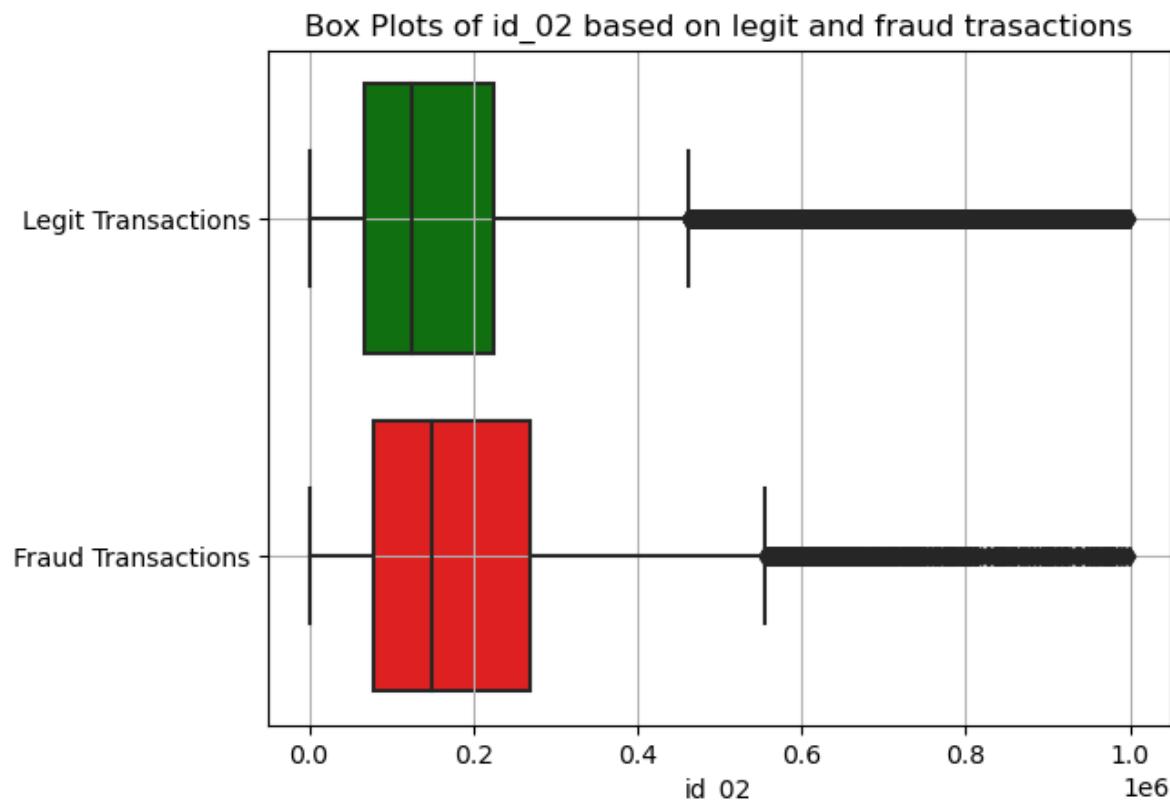
Percentile	Legit Transactions	Fraud Transactions
0	-100.0	-100.0
5	-35.0	-50.0
10	-20.0	-40.0
15	-20.0	-30.0
20	-15.0	-20.0
25	-10.0	-20.0
30	-5.0	-20.0
35	-5.0	-15.0
40	-5.0	-10.0
45	-5.0	-10.0
50	-5.0	-5.0
55	-5.0	-5.0
60	-5.0	-5.0
65	-5.0	-5.0
70	-5.0	-5.0
75	-5.0	-5.0
80	-5.0	-5.0
85	-5.0	-5.0
90	0.0	-5.0
95	0.0	0.0
100	0.0	0.0

InterQuartileRange of legit transactions: 5.0

InterQuartileRange of fraud transactions: 15.0



Analysis of `id_02` feature



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	140872.000000	count	129650.000000	
mean	174716.584708	mean	172396.362892	
std	159651.816856	std	158181.415938	
min	1.000000	min	2.000000	
25%	67992.000000	25%	67139.000000	
50%	79346.250000	50%	124125.000000	
75%	125800.500000	75%	225390.500000	
max	149599.500000	max	999595.000000	
		Name: id_02, dtype: float64	Name: id_02, dtype: float64	
	Name: id_02, dtype: float64			

The unique values in id_02 feature is: 115656

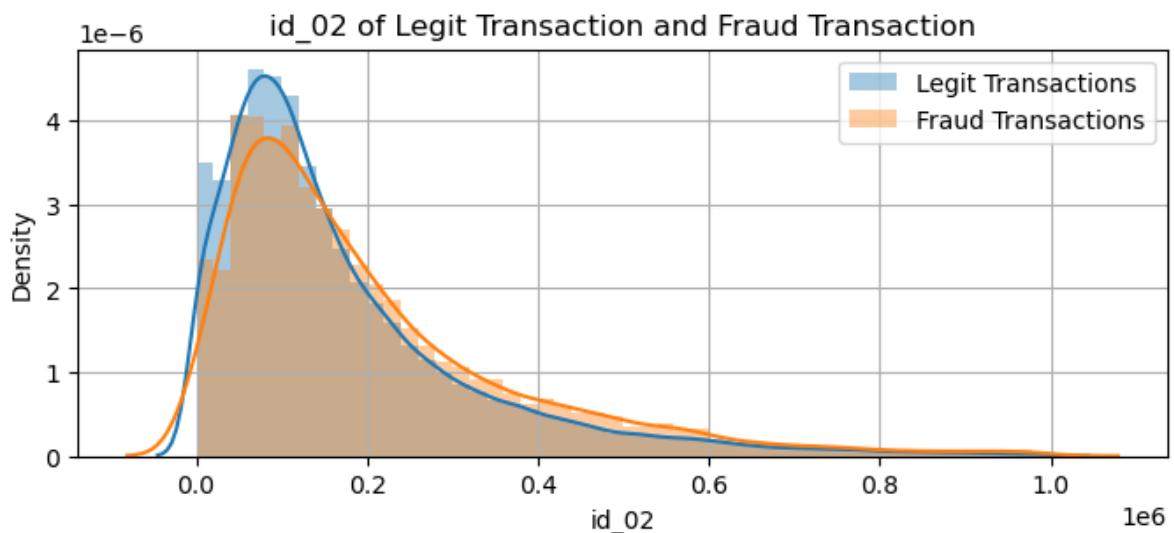
id_02 feature has 449668 null values, which holds 0.7614522301622244 %

Percentiles

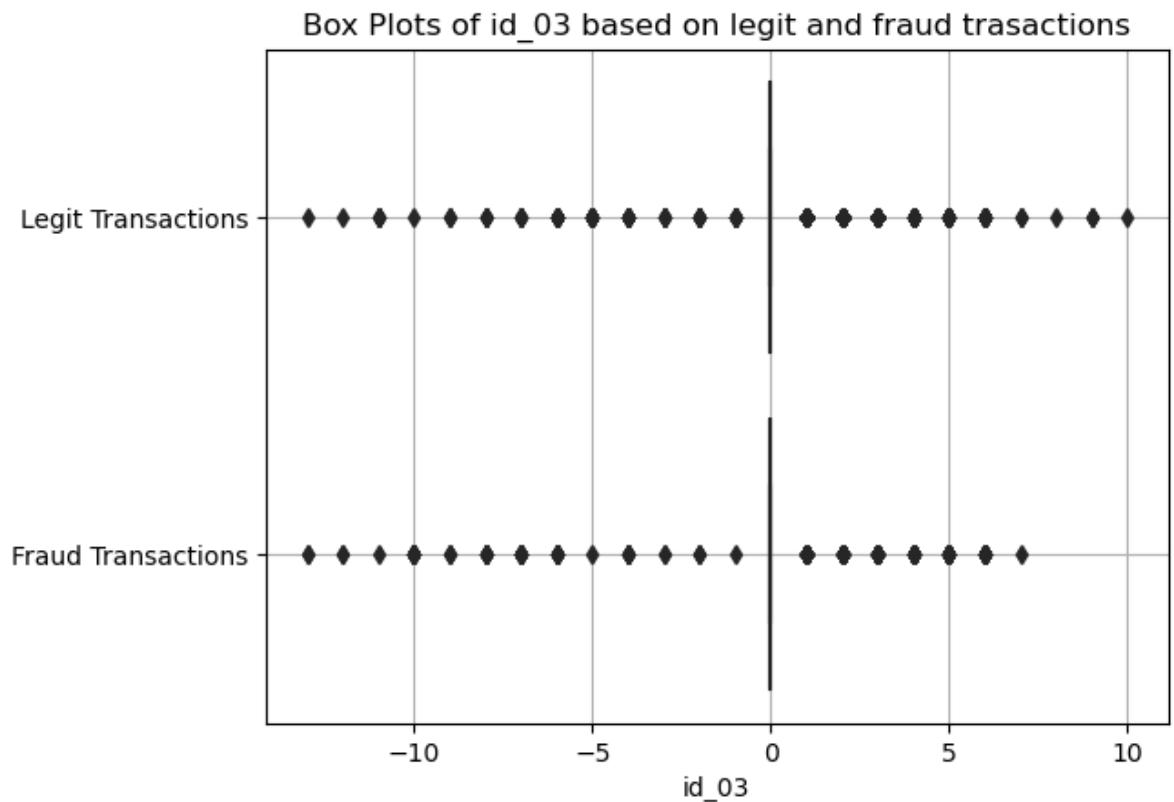
Percentile	Legit Transactions	Fraud Transactions
0	2.0	1.0
5	10818.15	23172.25
10	29724.0	42194.2
15	43750.4	54199.75
20	56231.8	67204.8
25	67139.0	79346.25
30	78051.7	92230.3
35	88819.15	105342.85
40	100065.6	117471.4
45	111480.1	133122.8
50	124125.0	149599.5
55	138592.95	167476.6
60	154866.4	186741.2
65	174214.35	209010.8
70	197596.6	235248.2
75	225390.5	269828.25
80	261155.2	310086.8
85	310091.2	367667.25
90	380950.2	445252.9
95	505642.05	560824.25
100	999595.0	998295.0

InterQuartileRange of legit transactions: 158251.5

InterQuartileRange of fraud transactions: 190482.0



Analysis of `id_03` feature



Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	66324.000000	count	59212.000000	
count	7112.000000			
mean	0.060189	mean	0.051594	
mean	0.131749			
std	0.598231	std	0.488841	
std	1.158604			
min	-13.000000	min	-13.000000	
min	-13.000000			
25%	0.000000	25%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
max	10.000000	max	10.000000	
max	7.000000			
Name: id_03, dtype: float64		Name: id_03, dtype: float64		
Name: id_03, dtype: float64				

The unique values in id_03 feature is: 25

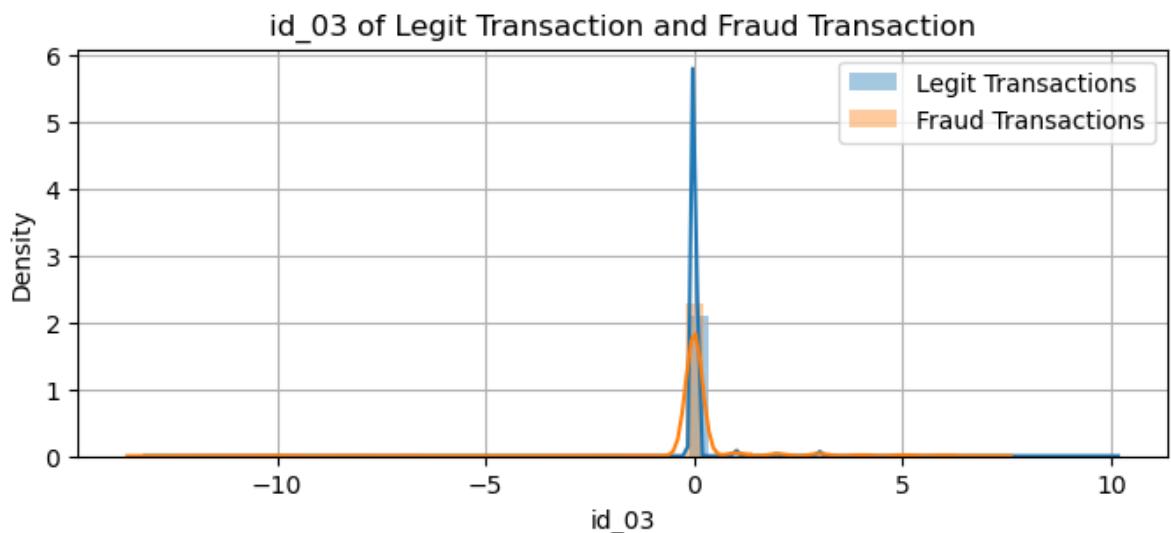
id_03 feature has 524216 null values, which holds 0.8876892335828225 %

Percentiles

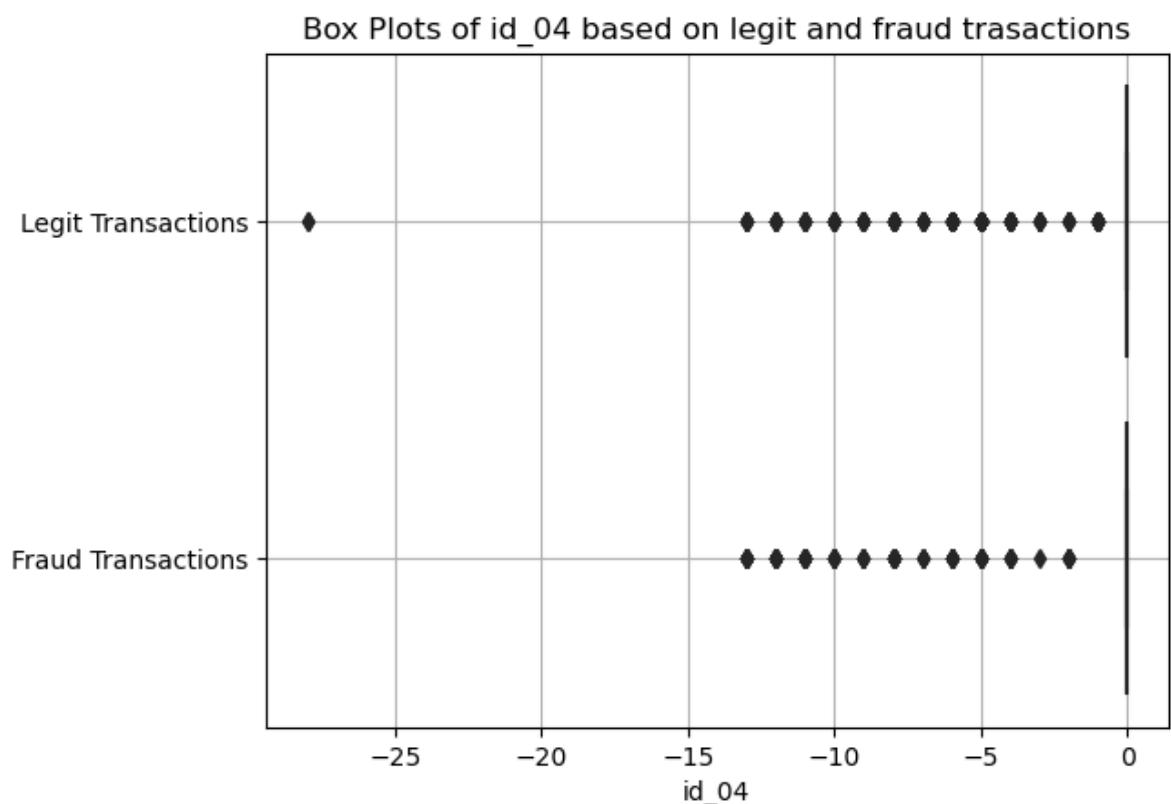
Percentile	Legit Transactions	Fraud Transactions
0	-13.0	-13.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	0.0
80	0.0	0.0
85	0.0	0.0
90	0.0	0.0
95	0.0	2.0
100	10.0	7.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 0.0



Analysis of `id_04` feature



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	66324.000000	count	59212.000000	
mean	-0.058938	mean	-0.044434	
std	0.701015	std	0.583593	
min	-28.000000	min	-28.000000	
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	0.000000	75%	0.000000	
max	0.000000	max	0.000000	
	Name: id_04, dtype: float64		Name: id_04, dtype: float64	
	Name: id_04, dtype: float64			

The unique values in id_04 feature is: 16

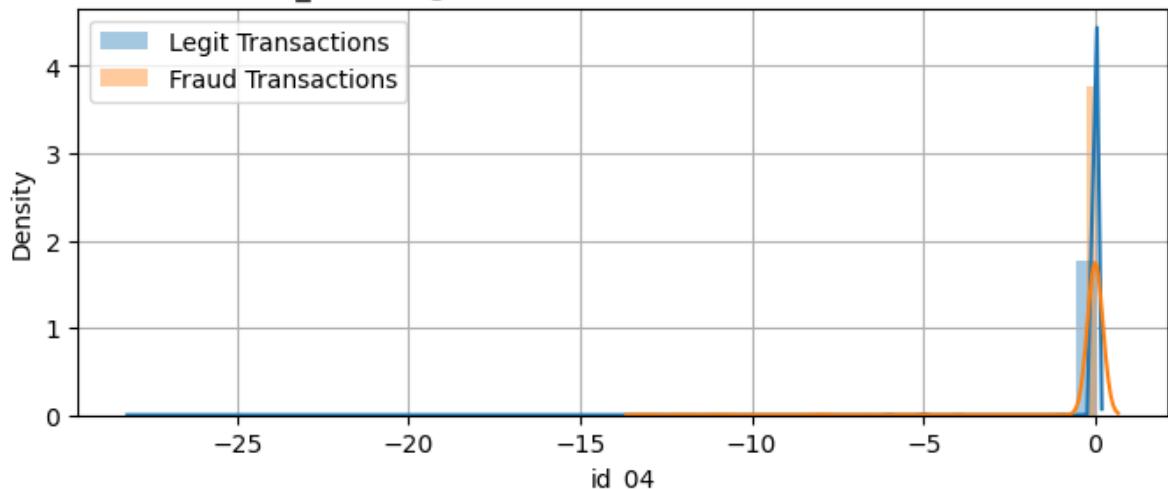
id_04 feature has 524216 null values, which holds 0.8876892335828225 %

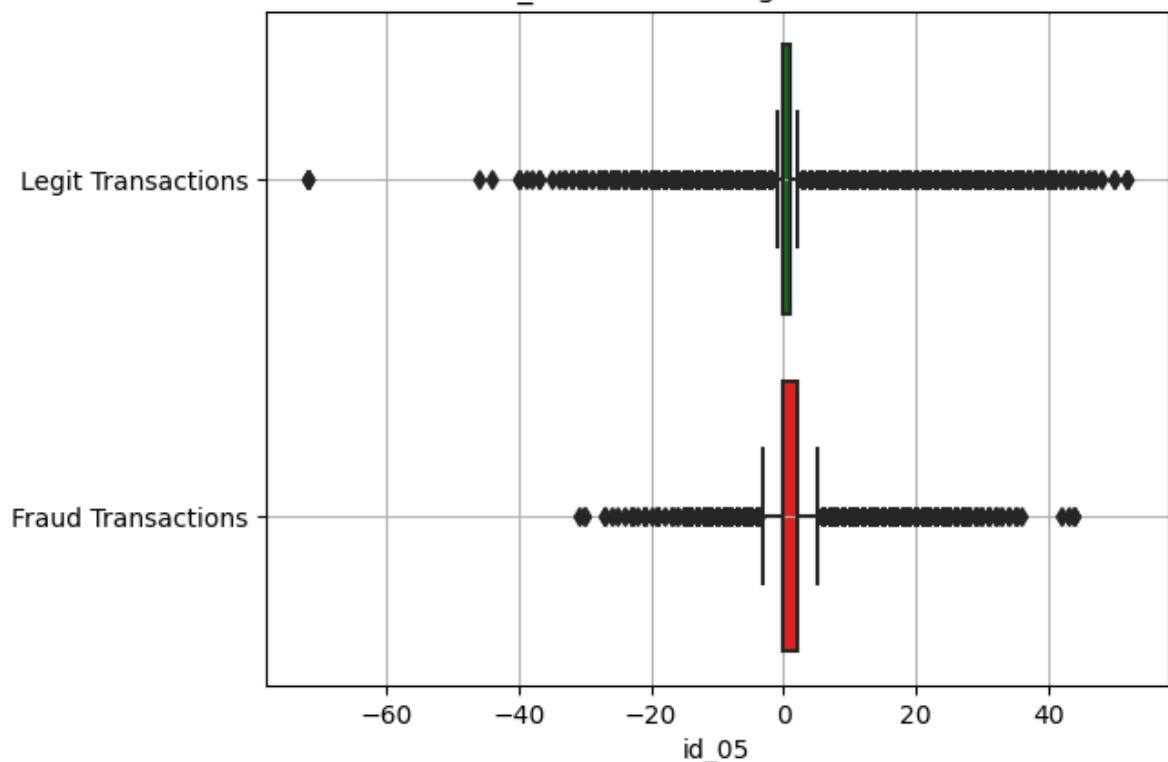
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	-28.0	-13.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	0.0
80	0.0	0.0
85	0.0	0.0
90	0.0	0.0
95	0.0	0.0
100	0.0	0.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 0.0

id_04 of Legit Transaction and Fraud Transaction

Analysis of id_05 feature**Box Plots of id_05 based on legit and fraud transactions**

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	136865.000000	count	125883.000000	
count	10982.000000			
mean	1.615585	mean	1.627956	
mean	1.473775			
std	5.249856	std	5.245558	
std	5.297058			
min	-72.000000	min	-72.000000	
min	-31.000000			
25%	0.000000	25%	0.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			
max	52.000000	max	52.000000	
max	44.000000			
Name: id_05, dtype: float64		Name: id_05, dtype: float64		
Name: id_05, dtype: float64				

The unique values in id_05 feature is: 94

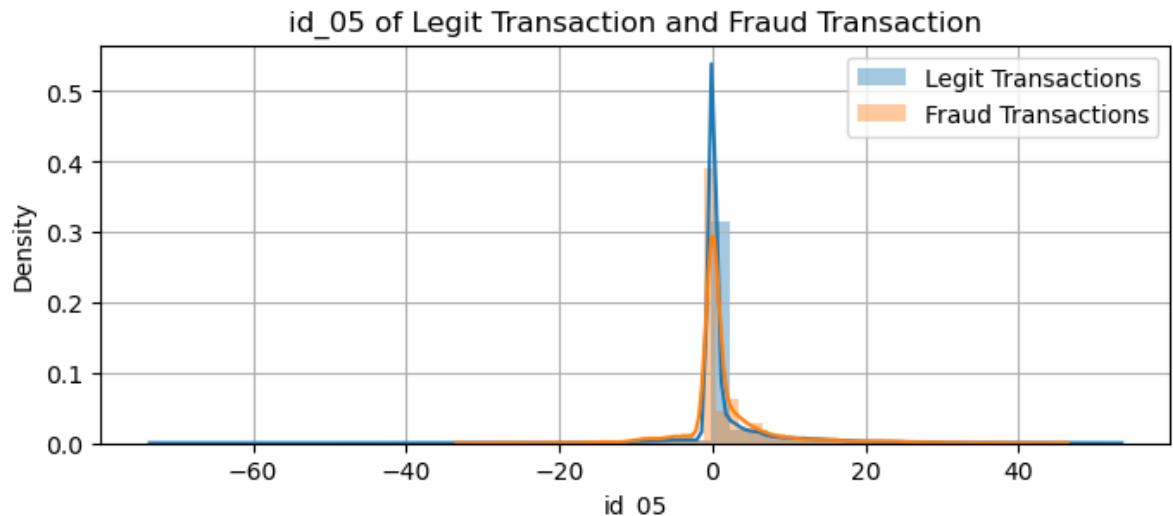
id_05 feature has 453675 null values, which holds 0.7682375452975243 %

Percentiles

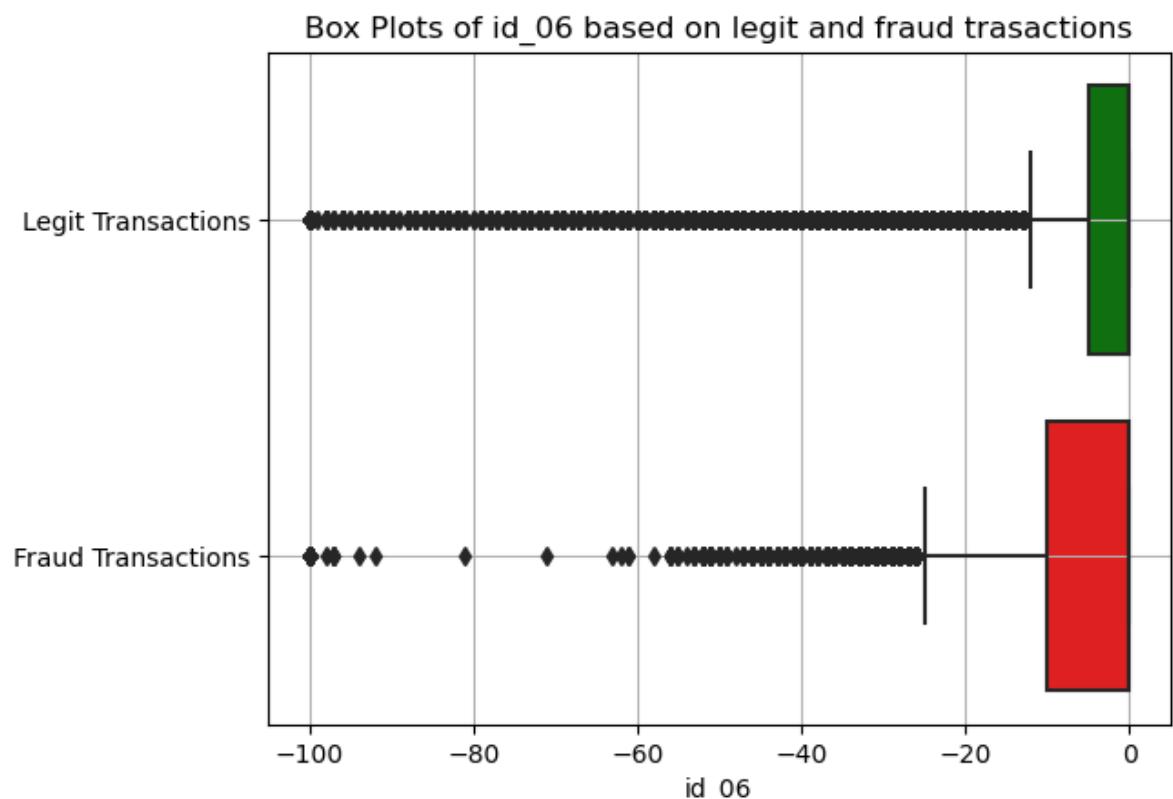
Percentile	Legit Transactions	Fraud Transactions
0	-72.0	-31.0
5	0.0	-5.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	1.0
75	1.0	2.0
80	2.0	3.0
85	4.0	4.0
90	6.0	6.0
95	12.0	12.0
100	52.0	44.0

InterQuartileRange of legit transactions: 1.0

InterQuartileRange of fraud transactions: 2.0



Analysis of id_06 feature



Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	136865.000000	count	125883.000000	
count	10982.000000			
mean	-6.698710	mean	-6.566518	
mean	-8.213987			
std	16.491104	std	16.442454	
std	16.966203			
min	-100.000000	min	-100.000000	
min	-100.000000			
25%	-6.000000	25%	-5.000000	
25%	-10.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	0.000000	max	0.000000	
max	0.000000			
Name: id_06, dtype: float64		Name: id_06, dtype: float64		
Name: id_06, dtype: float64				

The unique values in id_06 feature is: 102

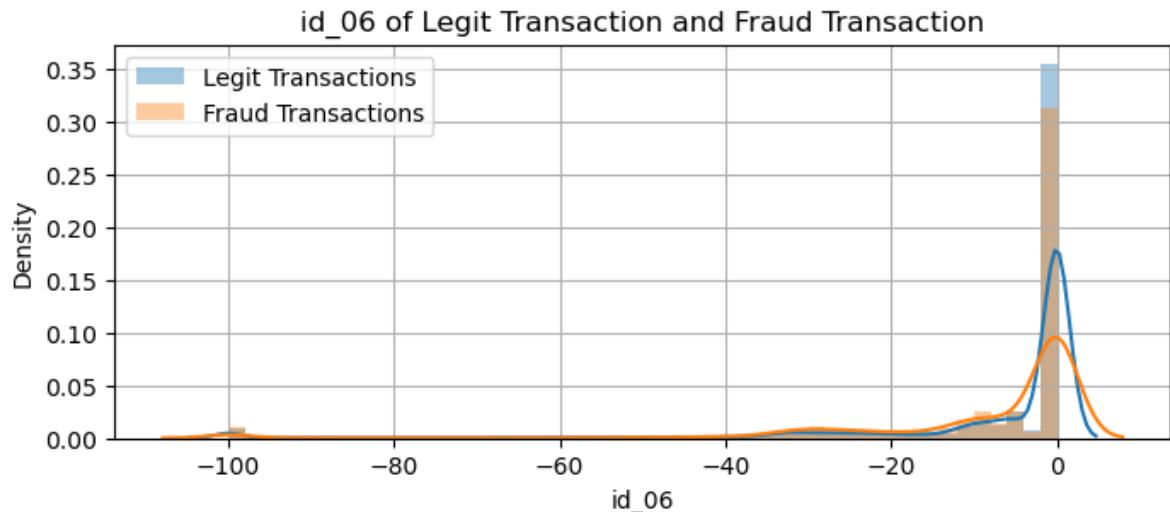
id_06 feature has 453675 null values, which holds 0.7682375452975243 %

Percentiles

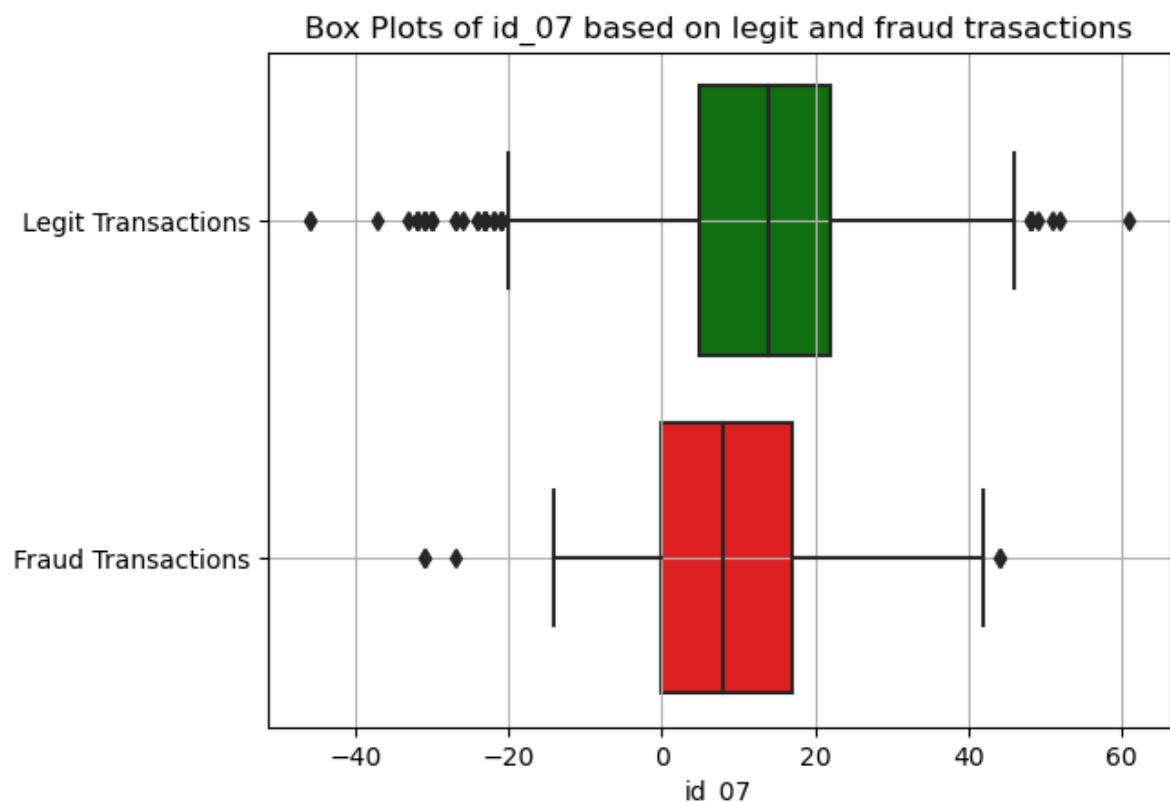
Percentile	Legit Transactions	Fraud Transactions
0	-100.0	-100.0
5	-33.0	-33.0
10	-24.0	-28.0
15	-12.0	-20.0
20	-8.0	-12.0
25	-5.0	-10.0
30	-1.0	-8.0
35	0.0	-5.0
40	0.0	-1.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	0.0
80	0.0	0.0
85	0.0	0.0
90	0.0	0.0
95	0.0	0.0
100	0.0	0.0

InterQuartileRange of legit transactions: 5.0

InterQuartileRange of fraud transactions: 10.0



Analysis of `id_07` feature



Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	5155.000000	count	4729.000000	
count	426.000000			
mean	13.285354	mean	13.574963	
mean	10.070423	std	11.248968	
std	11.384207	min	-46.000000	
std	12.355698	25%	5.000000	
min	-46.000000	50%	14.000000	
min	-31.000000	75%	22.000000	
25%	5.000000	max	61.000000	
50%	14.000000	max	61.000000	
50%	8.000000			
75%	22.000000			
75%	17.000000			
max	61.000000			
max	44.000000			
	Name: id_07, dtype: float64		Name: id_07, dtype: float64	
	Name: id_07, dtype: float64			

The unique values in id_07 feature is: 85

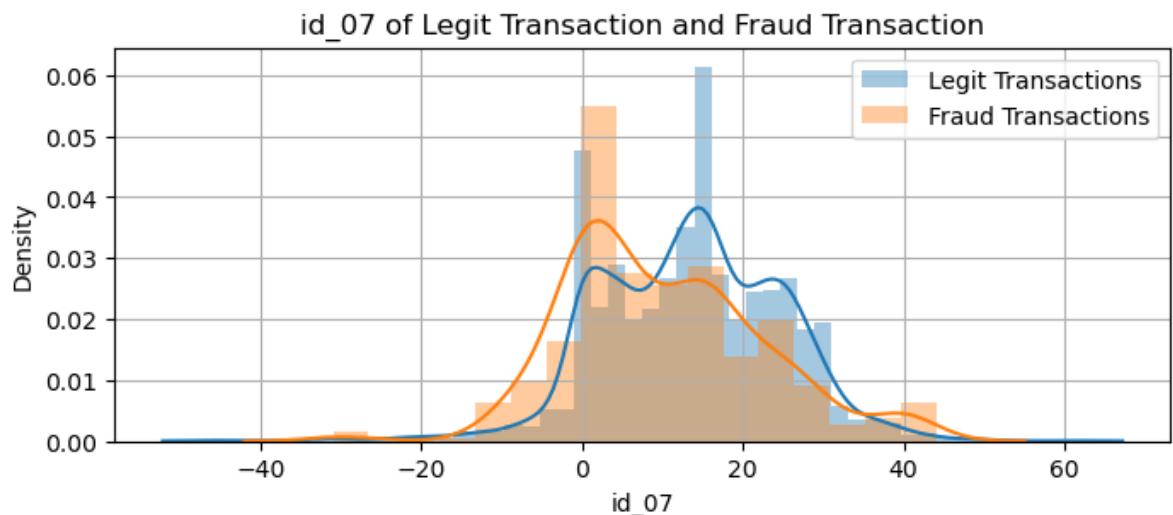
id_07 feature has 585385 null values, which holds 0.9912707013919464 %

Percentiles

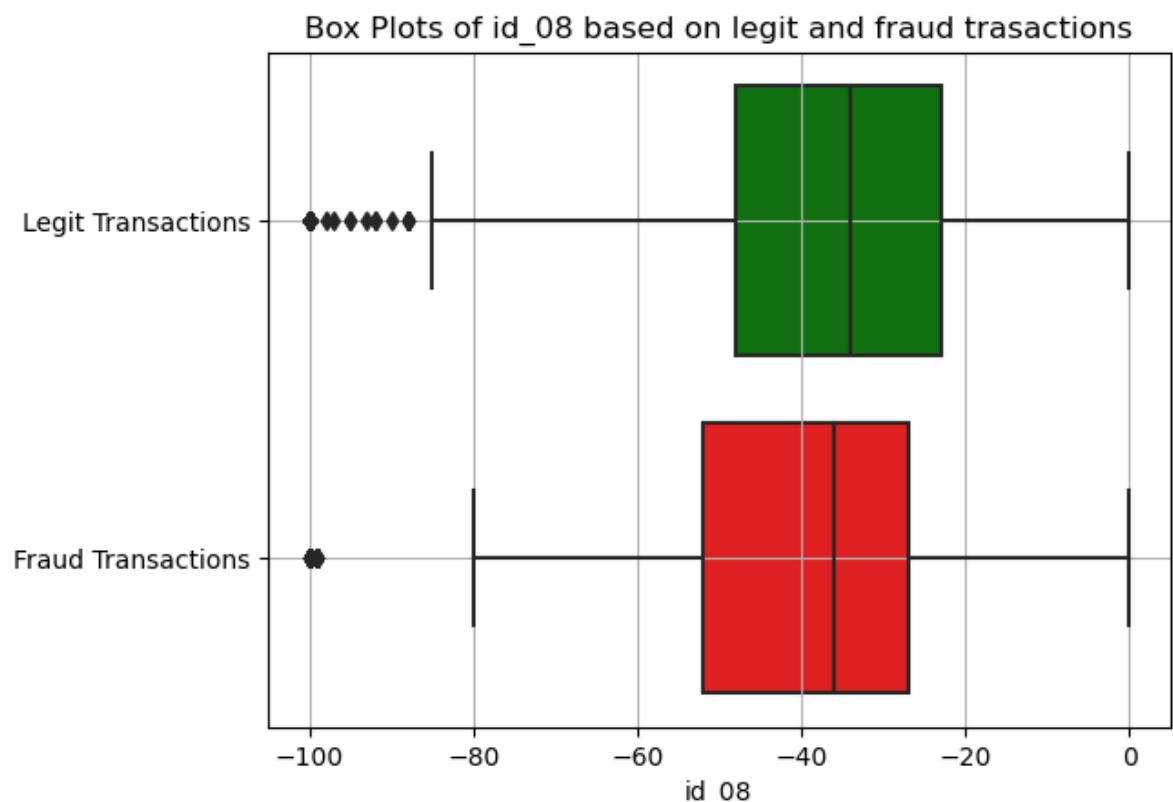
Percentile	Legit Transactions	Fraud Transactions
0	-46.0	-31.0
5	-2.0	-6.0
10	0.0	-3.0
15	1.0	-1.0
20	3.0	0.0
25	5.0	0.0
30	7.0	2.0
35	9.0	3.0
40	11.0	5.0
45	13.0	6.0
50	14.0	8.0
55	15.0	11.0
60	16.0	13.0
65	18.0	14.0
70	20.0	16.0
75	22.0	17.0
80	24.0	20.0
85	25.0	23.0
90	27.0	26.0
95	30.0	31.0
100	61.0	44.0

InterQuartileRange of legit transactions: 17.0

InterQuartileRange of fraud transactions: 17.0



Analysis of `id_08` feature



Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	5155.000000	count	4729.000000	
count	426.000000			
mean	-38.600388	mean	-38.150349	
mean	-43.596244			
std	26.084899	std	25.894971	
std	27.651042			
min	-100.000000	min	-100.000000	
min	-100.000000			
25%	-48.000000	25%	-48.000000	
25%	-52.000000			
50%	-34.000000	50%	-34.000000	
50%	-36.000000			
75%	-23.000000	75%	-23.000000	
75%	-27.000000			
max	0.000000	max	0.000000	
max	0.000000			
Name: id_08, dtype: float64		Name: id_08, dtype: float64		
Name: id_08, dtype: float64				

The unique values in id_08 feature is: 95

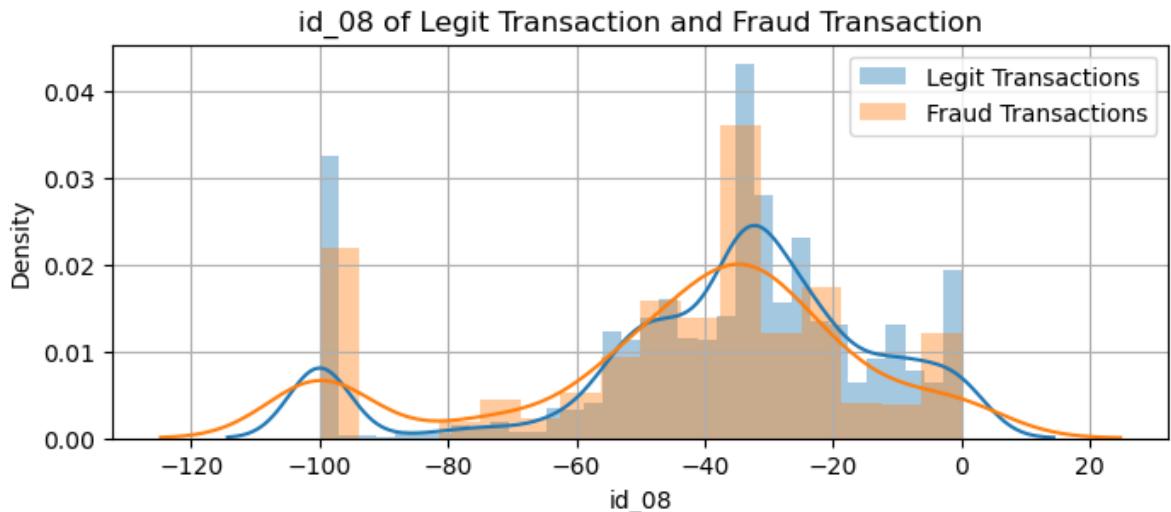
id_08 feature has 585385 null values, which holds 0.9912707013919464 %

Percentiles

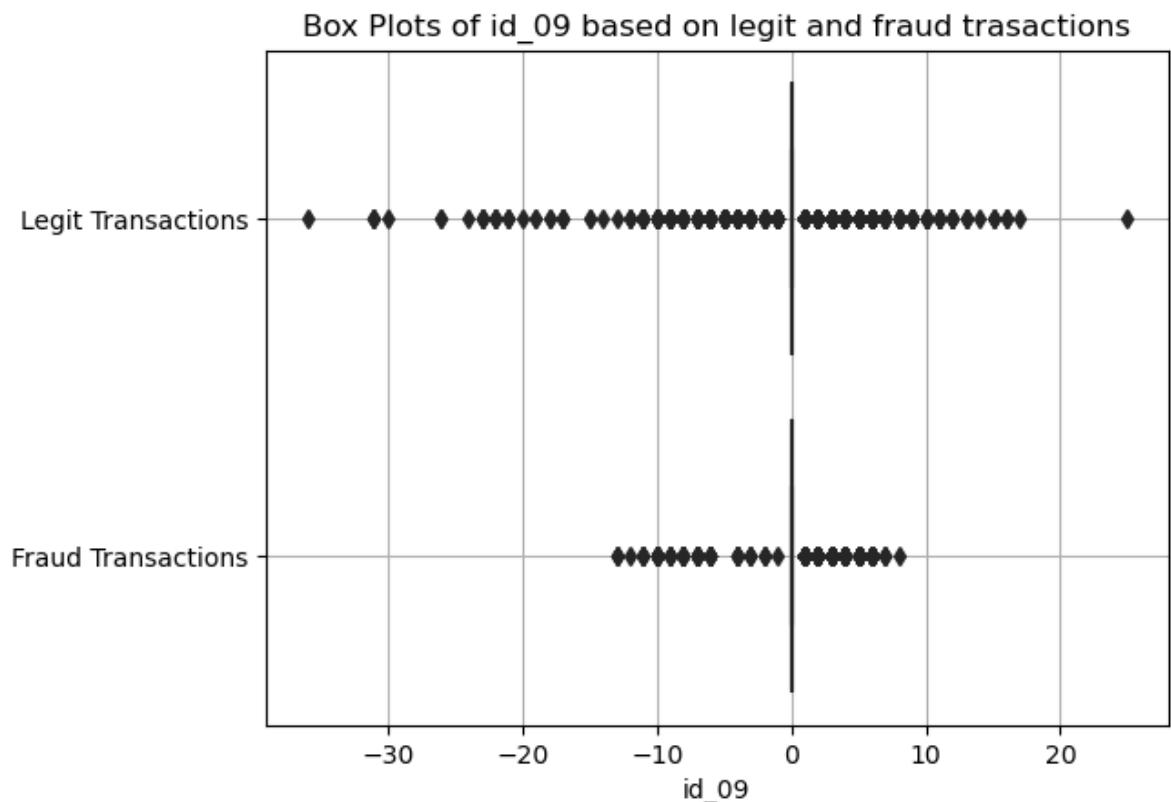
Percentile	Legit Transactions	Fraud Transactions
0	-100.0	-100.0
5	-100.0	-100.0
10	-82.0	-100.0
15	-56.0	-74.0
20	-52.0	-58.0
25	-48.0	-52.0
30	-45.0	-49.0
35	-41.0	-47.0
40	-36.8	-43.0
45	-35.0	-38.0
50	-34.0	-36.0
55	-32.0	-35.0
60	-31.0	-34.0
65	-28.0	-33.0
70	-25.0	-30.5
75	-23.0	-27.0
80	-19.0	-22.0
85	-13.0	-20.0
90	-9.0	-13.5
95	0.0	0.0
100	0.0	0.0

InterQuartileRange of legit transactions: 25.0

InterQuartileRange of fraud transactions: 25.0



Analysis of `id_09` feature



Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	74926.000000	count	67099.000000	
count	7827.000000			
mean	0.091023	mean	0.081134	
mean	0.175802			
std	0.983842	std	0.946010	
std	1.259368			
min	-36.000000	min	-36.000000	
min	-13.000000			
25%	0.000000	25%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
max	25.000000	max	25.000000	
max	8.000000			
	Name: id_09, dtype: float64		Name: id_09, dtype: float64	
	Name: id_09, dtype: float64			

The unique values in id_09 feature is: 47

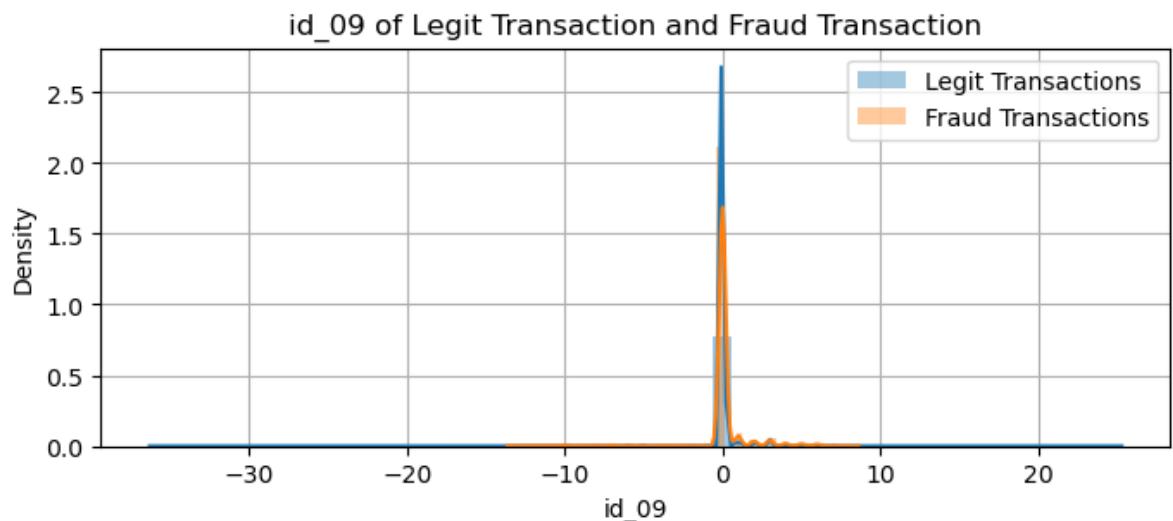
id_09 feature has 515614 null values, which holds 0.8731229044603245 %

Percentiles

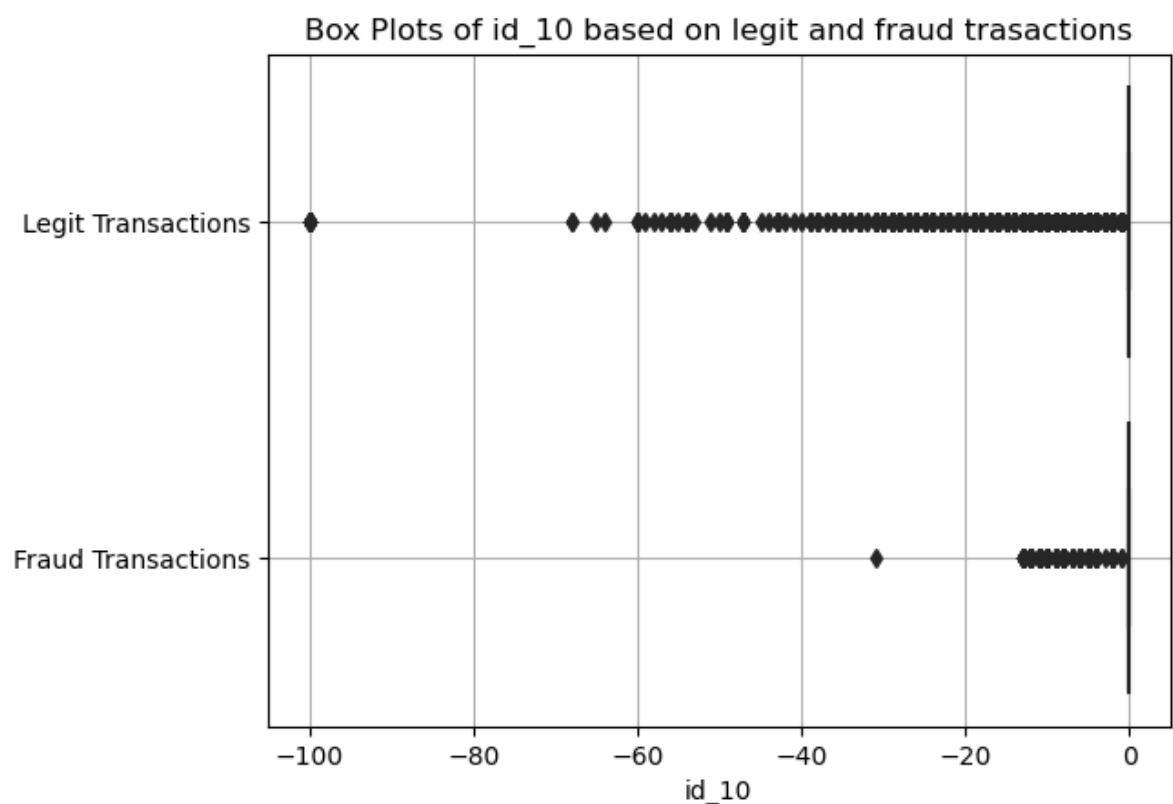
Percentile	Legit Transactions	Fraud Transactions
0	-36.0	-13.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	0.0
80	0.0	0.0
85	0.0	0.0
90	0.0	1.0
95	0.0	2.0
100	25.0	8.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 0.0



Analysis of id_10 feature



Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	74926.000000	count	67099.000000	
mean	-0.301124	mean	-0.311644	
std	2.789446	std	2.904755	
min	-100.000000	min	-100.000000	
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	0.000000	75%	0.000000	
max	0.000000	max	0.000000	
	Name: id_10, dtype: float64		Name: id_10, dtype: float64	
	Name: id_10, dtype: float64			

The unique values in id_10 feature is: 63

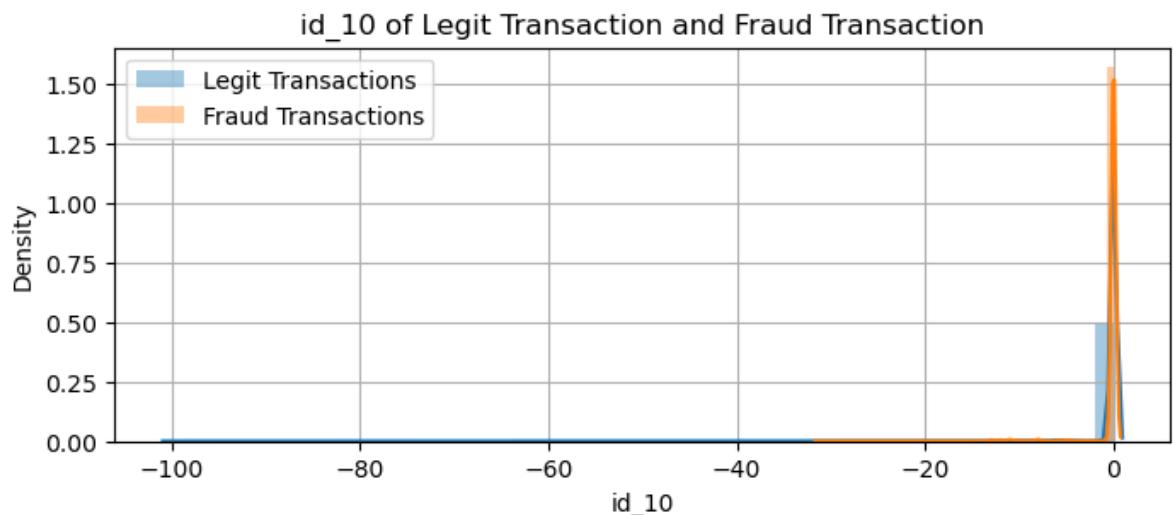
id_10 feature has 515614 null values, which holds 0.8731229044603245 %

Percentiles

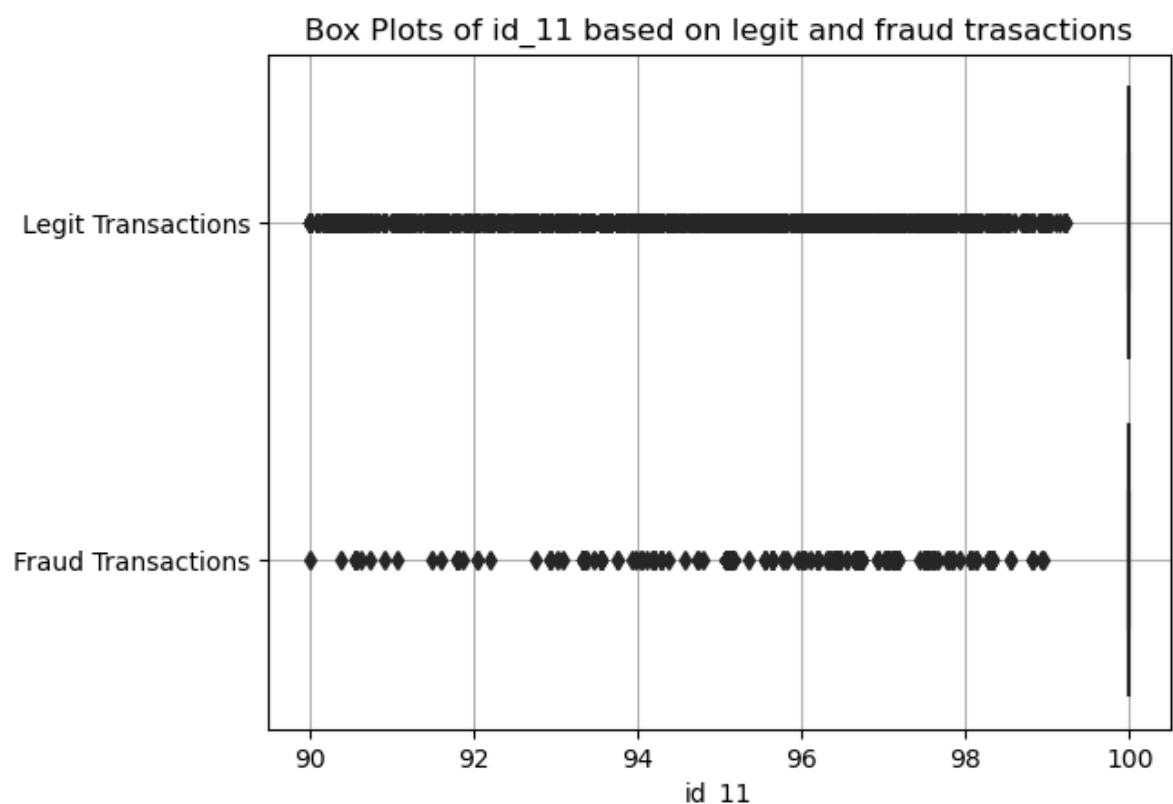
Percentile	Legit Transactions	Fraud Transactions
0	-100.0	-31.0
5	0.0	0.0
10	0.0	0.0
15	0.0	0.0
20	0.0	0.0
25	0.0	0.0
30	0.0	0.0
35	0.0	0.0
40	0.0	0.0
45	0.0	0.0
50	0.0	0.0
55	0.0	0.0
60	0.0	0.0
65	0.0	0.0
70	0.0	0.0
75	0.0	0.0
80	0.0	0.0
85	0.0	0.0
90	0.0	0.0
95	0.0	0.0
100	0.0	0.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 0.0



Analysis of `id_11` feature



Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	140978.000000	count	129760.000000	
count	11218.000000			
mean	99.745325	mean	99.742701	
mean	99.775677			
std	1.127602	std	1.137140	
std	1.010306			
min	90.000000	min	90.000000	
min	90.000000			
25%	100.000000	25%	100.000000	
50%	100.000000			
75%	100.000000	75%	100.000000	
max	100.000000	max	100.000000	
	100.000000			
Name: id_11, dtype: float64		Name: id_11, dtype: float64		
Name: id_11, dtype: float64				

The unique values in id_11 feature is: 366

id_11 feature has 449562 null values, which holds 0.7612727334304196 %

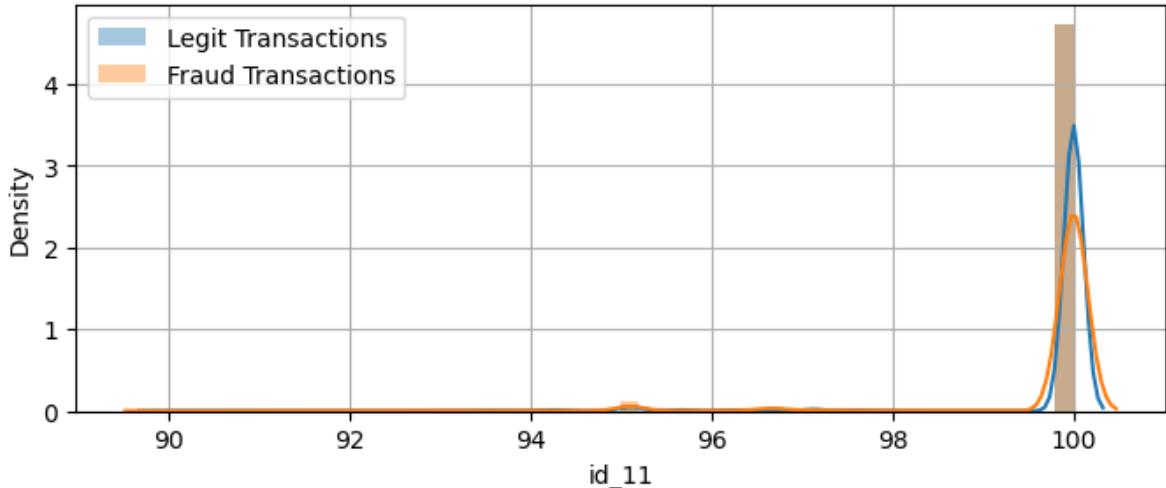
Percentiles

Percentile	Legit Transactions	Fraud Transactions
0	90.0	90.0
5	97.329	97.909
10	100.0	100.0
15	100.0	100.0
20	100.0	100.0
25	100.0	100.0
30	100.0	100.0
35	100.0	100.0
40	100.0	100.0
45	100.0	100.0
50	100.0	100.0
55	100.0	100.0
60	100.0	100.0
65	100.0	100.0
70	100.0	100.0
75	100.0	100.0
80	100.0	100.0
85	100.0	100.0
90	100.0	100.0
95	100.0	100.0
100	100.0	100.0

InterQuartileRange of legit transactions: 0.0

InterQuartileRange of fraud transactions: 0.0

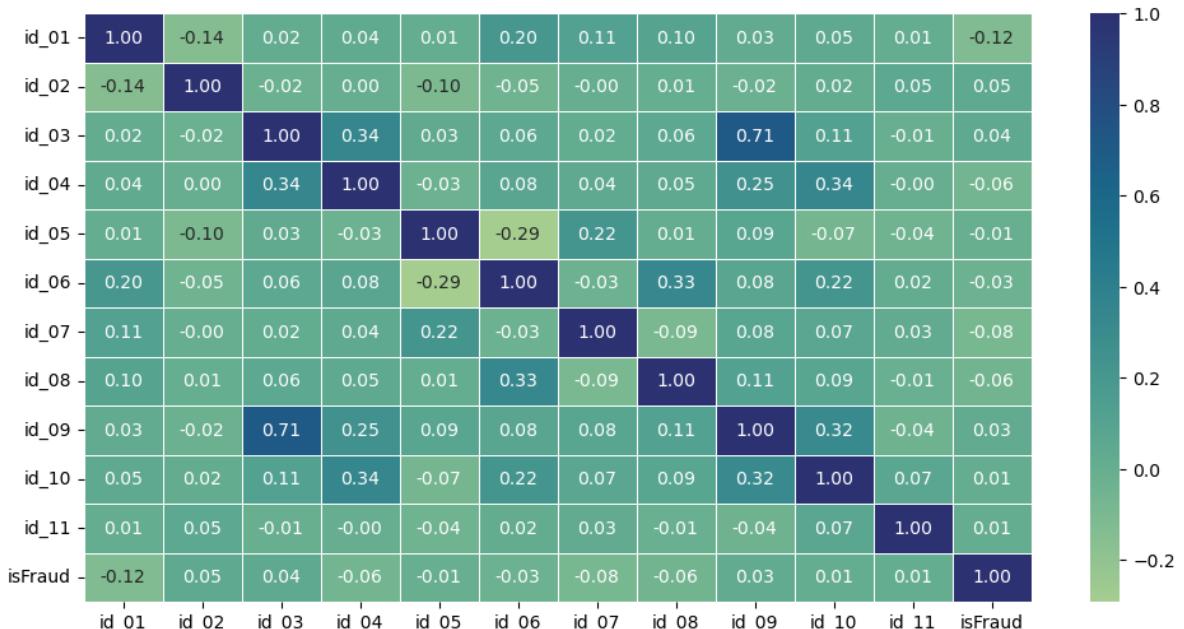
id_11 of Legit Transaction and Fraud Transaction



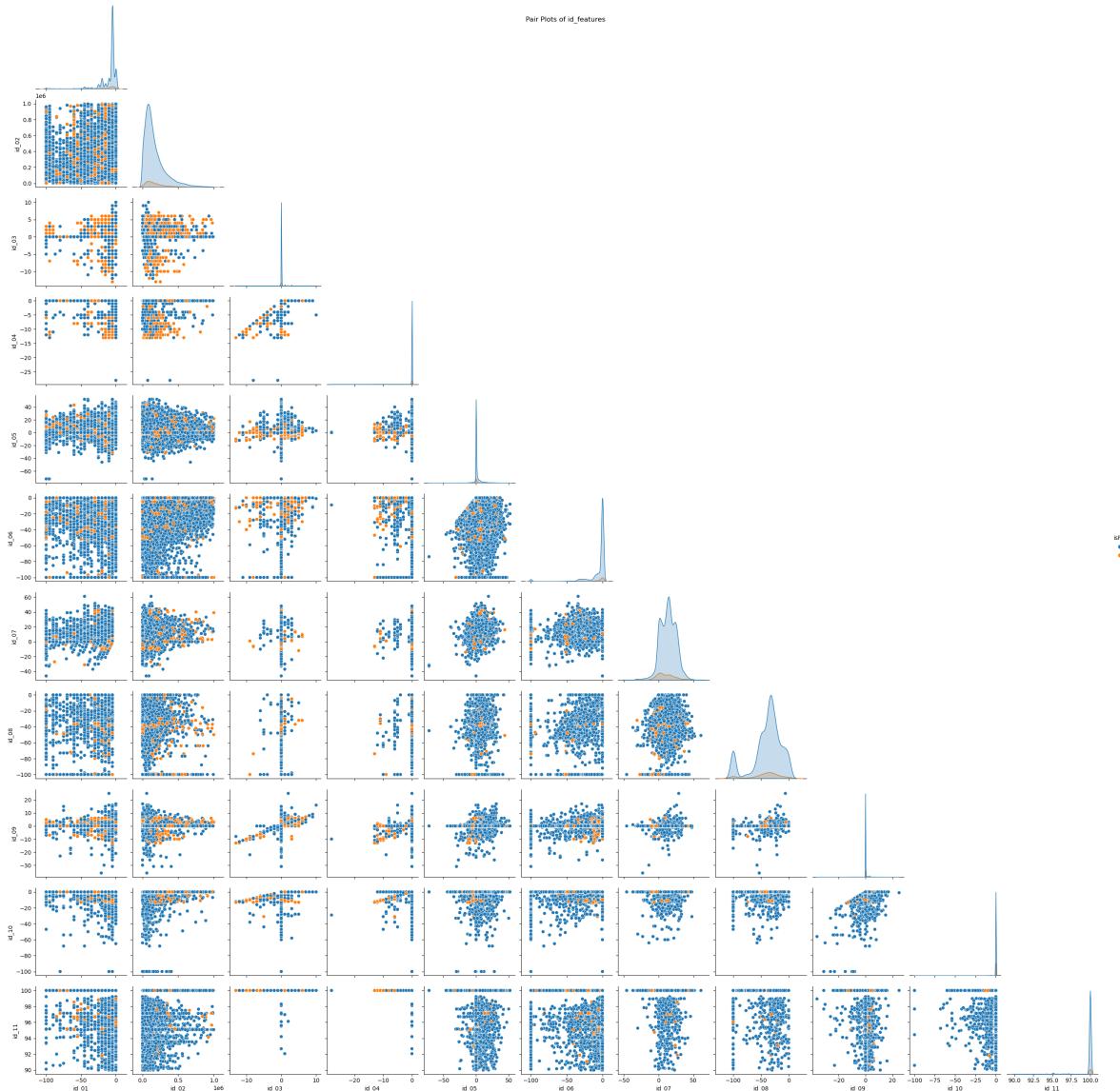
```
*****
*****
```

In [80]: `# to get collinearity of ID features`

```
plt.figure(figsize=(12,6))
sns.heatmap(df_train[id_features + ['isFraud']].corr(), annot = True, fmt=".2f", cbar_kws={"shrink": 0.5})
plt.show()
```



In [81]: `sns.pairplot(vars = id_features, hue='isFraud', data = df_train, corner=True)`
`plt.suptitle("\nPair Plots of id_features\n")`
`plt.show()`



Observation:

1. "id01 to id11 are numerical features for identity, which is collected by Vesta and security partners such as device rating, ip_domain rating, proxy rating, etc. Also it recorded behavioral fingerprint like account login times/failed to login times, how long an account stayed on the page, etc. All of these are not able to elaborate due to security partner T&C.
2. Based on the scatter plots, the id_feature seems important in identifying the transaction status.

1.2.7 Feature: 'V features'

In [82]: `def analysis_v_fea(data, feature, target):`

```

        data_legit = data[data[target]==0][feature]
        data_fraud = data[data[target]==1][feature]

        legit = pd.Series(data_legit)
        fraud = pd.Series(data_fraud)

        x = PrettyTable()
        x.field_names = ["Feature Description", "Feature Description of legit transact:
    
```

```
x.add_row([data[feature].describe(), legit.describe(), fraud.describe()])
print('\n')
print('*Five Number Summary*')
print(x)
print('\n')
print(f'The unique values in {feature} feature is:', len(data[feature].unique()))
print(f'{feature} feature has {data[feature].isnull().sum()} null values, which')
print('\n')
```

In [83]:

```
V_features = ['V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10', 'V11',
               'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20', 'V21',
               'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'V29', 'V30', 'V31',
               'V32', 'V33', 'V34', 'V35', 'V36', 'V37', 'V38', 'V39', 'V40', 'V41',
               'V42', 'V43', 'V44', 'V45', 'V46', 'V47', 'V48', 'V49', 'V50', 'V51',
               'V52', 'V53', 'V54', 'V55', 'V56', 'V57', 'V58', 'V59', 'V60', 'V61',
               'V62', 'V63', 'V64', 'V65', 'V66', 'V67', 'V68', 'V69', 'V70', 'V71',
               'V72', 'V73', 'V74', 'V75', 'V76', 'V77', 'V78', 'V79', 'V80', 'V81',
               'V82', 'V83', 'V84', 'V85', 'V86', 'V87', 'V88', 'V89', 'V90', 'V91',
               'V92', 'V93', 'V94', 'V95', 'V96', 'V97', 'V98', 'V99', 'V100', 'V101',
               'V102', 'V103', 'V104', 'V105', 'V106', 'V107', 'V108', 'V109', 'V110',
               'V111', 'V112', 'V113', 'V114', 'V115', 'V116', 'V117', 'V118', 'V119',
               'V120', 'V121', 'V122', 'V123', 'V124', 'V125', 'V126', 'V127', 'V128',
               'V129', 'V130', 'V131', 'V132', 'V133', 'V134', 'V135', 'V136', 'V137',
               'V138', 'V139', 'V140', 'V141', 'V142', 'V143', 'V144', 'V145', 'V146',
               'V147', 'V148', 'V149', 'V150', 'V151', 'V152', 'V153', 'V154', 'V155',
               'V156', 'V157', 'V158', 'V159', 'V160', 'V161', 'V162', 'V163', 'V164',
               'V165', 'V166', 'V167', 'V168', 'V169', 'V170', 'V171', 'V172', 'V173',
               'V174', 'V175', 'V176', 'V177', 'V178', 'V179', 'V180', 'V181', 'V182',
               'V183', 'V184', 'V185', 'V186', 'V187', 'V188', 'V189', 'V190', 'V191',
               'V192', 'V193', 'V194', 'V195', 'V196', 'V197', 'V198', 'V199', 'V200',
               'V201', 'V202', 'V203', 'V204', 'V205', 'V206', 'V207', 'V208', 'V209',
               'V210', 'V211', 'V212', 'V213', 'V214', 'V215', 'V216', 'V217', 'V218',
               'V219', 'V220', 'V221', 'V222', 'V223', 'V224', 'V225', 'V226', 'V227',
               'V228', 'V229', 'V230', 'V231', 'V232', 'V233', 'V234', 'V235', 'V236',
               'V237', 'V238', 'V239', 'V240', 'V241', 'V242', 'V243', 'V244', 'V245',
               'V246', 'V247', 'V248', 'V249', 'V250', 'V251', 'V252', 'V253', 'V254',
               'V255', 'V256', 'V257', 'V258', 'V259', 'V260', 'V261', 'V262', 'V263',
               'V264', 'V265', 'V266', 'V267', 'V268', 'V269', 'V270', 'V271', 'V272',
               'V273', 'V274', 'V275', 'V276', 'V277', 'V278', 'V279', 'V280', 'V281',
               'V282', 'V283', 'V284', 'V285', 'V286', 'V287', 'V288', 'V289', 'V290',
               'V291', 'V292', 'V293', 'V294', 'V295', 'V296', 'V297', 'V298', 'V299',
               'V300', 'V301', 'V302', 'V303', 'V304', 'V305', 'V306', 'V307', 'V308',
               'V309', 'V310', 'V311', 'V312', 'V313', 'V314', 'V315', 'V316', 'V317',
               'V318', 'V319', 'V320', 'V321', 'V322', 'V323', 'V324', 'V325', 'V326',
               'V327', 'V328', 'V329', 'V330', 'V331', 'V332', 'V333', 'V334', 'V335',
               'V336', 'V337', 'V338', 'V339']
```

```
for fea in V_features:
    print(f'Analysis of {fea} feature')
    analysis_v_fea(df_train, fea, 'isFraud')
    print('\n')
    print('*'*120)
    print('\n')
```

Analysis of V1 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	311253.000000	count	305147.000000	
count	6106.0			
mean	0.999945	mean	0.999944	
mean	1.0			
std	0.007390	std	0.007464	
std	0.0			
min	0.000000	min	0.000000	
min	1.0			
25%	1.000000	25%	1.000000	
25%	1.0			
50%	1.000000	50%	1.000000	
50%	1.0			
75%	1.000000	75%	1.000000	
75%	1.0			
max	1.000000	max	1.000000	
max	1.0			
	Name: V1, dtype: float64		Name: V1, dtype: float64	
	Name: V1, dtype: float64			

The unique values in V1 feature is: 3

V1 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V2 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	311253.000000	count	305147.000000	
count	6106.000000			
mean	1.045204	mean	1.044529	
mean	1.078939			
std	0.240133	std	0.237326	
std	0.351422			
min	0.000000	min	0.000000	
min	1.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	1.000000	75%	1.000000
75%	1.000000		
max	8.000000	max	8.000000
max	6.000000		
Name: V2, dtype: float64		Name: V2, dtype: float64	
Name: V2, dtype: float64			

The unique values in V2 feature is: 10

V2 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V3 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	311253.000000	count	305147.000000	
count	6106.000000			
mean	1.078075	mean	1.075839	
mean	1.189813			
std	0.320890	std	0.314484	
std	0.541936			
min	0.000000	min	0.000000	
min	1.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	9.000000	max	9.000000	
max	7.000000			
Name: V3, dtype: float64		Name: V3, dtype: float64		
Name: V3, dtype: float64				

The unique values in V3 feature is: 11

V3 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V4 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	311253.000000	count	305147.000000	count
mean	0.846456	mean	0.845671	mean
std	0.440053	std	0.438623	std
min	0.000000	min	0.000000	min
25%	1.000000	25%	1.000000	25%
50%	1.000000	50%	1.000000	50%
75%	1.000000	75%	1.000000	75%
max	6.000000	max	6.000000	max
	Name: V4, dtype: float64		Name: V4, dtype: float64	
	Name: V4, dtype: float64			

The unique values in V4 feature is: 8

V4 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V5 feature***Five Number Summary***

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	311253.000000	count	305147.000000	count
mean	0.876991	mean	0.874506	mean
std	0.475902	std	0.471002	std
min	0.000000	min	0.000000	min
25%	1.000000	25%	1.000000	25%
50%	1.000000	50%	1.000000	50%
75%	1.000000	75%	1.000000	75%
75%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	6.000000	max	6.000000
max	6.000000		
Name: V5, dtype: float64		Name: V5, dtype: float64	
Name: V5, dtype: float64			

The unique values in V5 feature is: 8

V5 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V6 feature

Five Number Summary

Feature Description of legit transactions		Feature Description of fraud transactions		Feature D
Feature Description of legit transactions		Feature Description of fraud transactions		cription of legit transactions Feature D
count	311253.000000	count	305147.000000	
count	6106.000000			
mean	1.045686	mean	1.045329	
mean	1.063544			
std	0.239385	std	0.237727	
std	0.310700			
min	0.000000	min	0.000000	
min	1.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	9.000000	max	9.000000	
max	6.000000			
Name: V6, dtype: float64		Name: V6, dtype: float64		
Name: V6, dtype: float64				

The unique values in V6 feature is: 11

V6 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V7 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	311253.000000	count	305147.000000	count
mean	1.072870	mean	1.071857	mean
std	0.304779	std	0.302124	std
min	0.000000	min	0.000000	min
25%	1.000000	25%	1.000000	25%
50%	1.000000	50%	1.000000	50%
75%	1.000000	75%	1.000000	75%
max	9.000000	max	9.000000	max
	Name: V7, dtype: float64		Name: V7, dtype: float64	
	Name: V7, dtype: float64			

The unique values in V7 feature is: 11

V7 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V8 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	311253.000000	count	305147.000000	count
mean	1.027704	mean	1.027282	mean
std	0.186069	std	0.183936	std
min	0.000000	min	0.000000	min
25%	1.000000	25%	1.000000	25%
50%	1.000000	50%	1.000000	50%
75%	1.000000	75%	1.000000	75%
max	8.000000	max	8.000000	max
	5.000000			

Name: V8, dtype: float64	Name: V8, dtype: float64
Name: V8, dtype: float64	
+-----+-----+	+-----+-----+
-----+-----+	

The unique values in V8 feature is: 10

V8 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V9 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	311253.000000	count	305147.000000	
count	6106.000000			
mean	1.041529	mean	1.040315	
mean	1.102195			
std	0.226339	std	0.222790	
std	0.356576			
min	0.000000	min	0.000000	
min	1.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	8.000000	max	8.000000	
max	5.000000			
Name: V9, dtype: float64	Name: V9, dtype: float64			
Name: V9, dtype: float64				
+-----+-----+	+-----+-----+			
-----+-----+				

The unique values in V9 feature is: 10

V9 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V10 feature

Five Number Summary

+-----+-----+	+-----+-----+	+-----+-----+
-----+-----+		

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	311253.000000	count	305147.000000
count	6106.000000		
mean	0.463915	mean	0.469423
mean	0.188667		
std	0.521522	std	0.521856
std	0.421117		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	1.000000	75%	1.000000
75%	0.000000		
max	4.000000	max	4.000000
max	4.000000		
Name: V10, dtype: float64		Name: V10, dtype: float64	
Name: V10, dtype: float64			

The unique values in V10 feature is: 6

V10 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V11 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	311253.000000	count	305147.000000
count	6106.000000		
mean	0.478987	mean	0.484609
mean	0.198002		
std	0.552431	std	0.552870
std	0.447692		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	1.000000	75%	1.000000
75%	0.000000		
max	5.000000	max	5.000000
max	4.000000		
Name: V11, dtype: float64		Name: V11, dtype: float64	
Name: V11, dtype: float64			

+-----+	+-----+	+-----+
- - - - -	- - - - -	- - - - -

The unique values in V11 feature is: 7
V11 feature has 279287 null values, which holds 0.47293494090154775 %

Analysis of V28 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	514467.000000	count	497691.000000	
count	16776.0			
mean	0.000830	mean	0.000858	
mean	0.0			
std	0.031133	std	0.031653	
std	0.0			
min	0.000000	min	0.000000	
min	0.0			
25%	0.000000	25%	0.000000	
25%	0.0			
50%	0.000000	50%	0.000000	
50%	0.0			
75%	0.000000	75%	0.000000	
75%	0.0			
max	4.000000	max	4.000000	
max	0.0			
Name: V28, dtype: float64		Name: V28, dtype: float64		
Name: V28, dtype: float64				

The unique values in V28 feature is: 5
V28 feature has 76073 null values, which holds 0.1288193856470349 %

Analysis of V29 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	514467.000000	count	497691.000000	
count	16776.0			

count	514467.000000		count	497691.000000
count	16776.000000		mean	0.397952
mean	0.387840		std	0.513065
mean	0.087864		min	0.000000
std	0.510652		25%	0.000000
std	0.307336		50%	0.000000
min	0.000000		75%	1.000000
min	0.000000		max	5.000000
25%	0.000000		max	5.000000
50%	0.000000			
75%	1.000000			
75%	0.000000			
max	5.000000			
max	4.000000			
	Name: V29, dtype: float64		Name: V29, dtype: float64	
	Name: V29, dtype: float64			

The unique values in V29 feature is: 7

V29 feature has 76073 null values, which holds 0.1288193856470349 %

Analysis of V30 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	514467.000000		count	497691.000000
count	16776.000000		mean	0.416927
mean	0.406436		std	0.557192
mean	0.095196		min	0.000000
std	0.554311		25%	0.000000
std	0.334857		50%	0.000000
min	0.000000		75%	1.000000
min	0.000000		max	9.000000
25%	0.000000		max	9.000000
25%	0.000000			
50%	0.000000			
75%	1.000000			
75%	0.000000			
max	9.000000			
max	5.000000			
	Name: V30, dtype: float64		Name: V30, dtype: float64	
	Name: V30, dtype: float64			

The unique values in V30 feature is: 9
V30 feature has 76073 null values, which holds 0.1288193856470349 %

Analysis of V31 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	514467.000000	count	497691.000000	
count	16776.000000			
mean	0.140761	mean	0.130103	
mean	0.456962			
std	0.355793	std	0.344070	
std	0.516374			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	5.000000			
Name: V31, dtype: float64		Name: V31, dtype: float64		
Name: V31, dtype: float64				

The unique values in V31 feature is: 9
V31 feature has 76073 null values, which holds 0.1288193856470349 %

Analysis of V32 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	514467.000000	count	497691.000000	
count	16776.000000			
mean	0.140761	mean	0.130103	
mean	0.456962			
std	0.355793	std	0.344070	
std	0.516374			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	5.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

count	514467.000000	count	497691.000000
count	16776.000000		
mean	0.142417	mean	0.131586
mean	0.463758		
std	0.368278	std	0.356108
std	0.538916		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	1.000000		
max	15.000000	max	15.000000
max	6.000000		
Name: V32, dtype: float64		Name: V32, dtype: float64	
Name: V32, dtype: float64			

The unique values in V32 feature is: 16

V32 feature has 76073 null values, which holds 0.1288193856470349 %

Analysis of V33 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	514467.000000	count	497691.000000	
count	16776.000000			
mean	0.130693	mean	0.119209	
mean	0.471388			
std	0.340900	std	0.327150	
std	0.518410			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	5.000000			
Name: V33, dtype: float64		Name: V33, dtype: float64		
Name: V33, dtype: float64				

The unique values in V33 feature is: 8
 V33 feature has 76073 null values, which holds 0.1288193856470349 %

Analysis of V34 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	514467.000000	count	497691.000000	
count	167776.000000			
mean	0.139154	mean	0.127318	
mean	0.490284			
std	0.357549	std	0.343634	
std	0.538387			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	13.000000	max	13.000000	
max	7.000000			
Name: V34, dtype: float64		Name: V34, dtype: float64		
Name: V34, dtype: float64				

The unique values in V34 feature is: 14
 V34 feature has 76073 null values, which holds 0.1288193856470349 %

Analysis of V35 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	0.542594	mean	0.550122
mean	0.332258	std	0.515322
std	0.516010	std	0.490180
min	0.000000	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	25%	0.000000
50%	1.000000	50%	0.000000
50%	0.000000	75%	1.000000
75%	1.000000	75%	1.000000
max	3.000000	max	3.000000
max	3.000000	Name: V35, dtype: float64	Name: V35, dtype: float64
		Name: V35, dtype: float64	

The unique values in V35 feature is: 5

V35 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V36 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.579198	mean	0.586323	
mean	0.380106	std	0.536893	
std	0.539039			
std	0.560396			
min	0.000000	min	0.000000	
min	0.000000	25%	0.000000	
25%	0.000000	25%	0.000000	
50%	1.000000	50%	1.000000	
50%	0.000000	75%	1.000000	
75%	1.000000	75%	1.000000	
max	5.000000	max	5.000000	
max	4.000000			
Name: V36, dtype: float64		Name: V36, dtype: float64		
Name: V36, dtype: float64				

The unique values in V36 feature is: 7

V36 feature has 168969 null values, which holds 0.2861262573238053 %

```
*****
*****
```

Analysis of V37 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	1.108065	mean	1.084948	
mean	1.753964			
std	0.690571	std	0.480715	
std	2.629251			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			
max	54.000000	max	19.000000	
max	54.000000			
Name: V37, dtype: float64		Name: V37, dtype: float64		
Name: V37, dtype: float64				

The unique values in V37 feature is: 56

V37 feature has 168969 null values, which holds 0.2861262573238053 %

```
*****
*****
```

Analysis of V38 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	1.16240	mean	1.130063	
mean	2.065902			

1. self_case_study_1_IEEE_fraud_detection_EDA

std	0.85892	std	0.605612
std	3.202819	min	0.000000
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
25%	1.000000	50%	1.000000
50%	1.000000	75%	1.000000
75%	2.000000	75%	1.000000
max	54.000000	max	27.000000
max	54.000000	Name: V38, dtype: float64	Name: V38, dtype: float64
Name: V38, dtype: float64		Name: V38, dtype: float64	

The unique values in V38 feature is: 56

V38 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V39 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.166076	mean	0.148711	
mean	0.651267			
std	0.451956	std	0.413146	
std	0.947924			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	15.000000	max	15.000000	
max	14.000000			
Name: V39, dtype: float64		Name: V39, dtype: float64		
Name: V39, dtype: float64				

The unique values in V39 feature is: 17

V39 feature has 168969 null values, which holds 0.2861262573238053 %

```
*****
*****
```

Analysis of V40 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.177145	mean	0.156817	
mean	0.745109			
std	0.505786	std	0.453107	
std	1.154619			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	1.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	24.000000	max	15.000000	
max	24.000000			
Name: V40, dtype: float64		Name: V40, dtype: float64		
Name: V40, dtype: float64				

The unique values in V40 feature is: 19

V40 feature has 168969 null values, which holds 0.2861262573238053 %

```
*****
*****
```

Analysis of V41 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.999269	mean	0.999248	
mean	0.999863			
std	0.027020	std	0.027409	
std	0.011717			

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000		min	0.000000	
min	0.000000				
25%	1.000000		25%	1.000000	
25%	1.000000				
50%	1.000000		50%	1.000000	
50%	1.000000				
75%	1.000000		75%	1.000000	
75%	1.000000				
max	1.000000		max	1.000000	
max	1.000000				
Name: V41, dtype: float64		Name: V41, dtype: float64			
Name: V41, dtype: float64					
+-----+-----+-----+					

The unique values in V41 feature is: 3

V41 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V42 feature

Five Number Summary

	Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	421571.000000		count	407004.000000		
count	14567.000000					
mean	0.156118		mean	0.142397		
mean	0.539507					
std	0.382896		std	0.365600		
std	0.596745					
min	0.000000		min	0.000000		
min	0.000000					
25%	0.000000		25%	0.000000		
25%	0.000000					
50%	0.000000		50%	0.000000		
50%	0.000000					
75%	0.000000		75%	0.000000		
75%	1.000000					
max	8.000000		max	8.000000		
max	5.000000					
Name: V42, dtype: float64		Name: V42, dtype: float64				
Name: V42, dtype: float64						
+-----+-----+-----+						

The unique values in V42 feature is: 10

V42 feature has 168969 null values, which holds 0.2861262573238053 %

```
*****
*****
```

Analysis of V43 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.168942	mean	0.152686	
mean	0.623121			
std	0.433359	std	0.407024	
std	0.769766			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	1.000000	75%	0.000000	
75%	1.000000			
max	8.000000	max	8.000000	
max	7.000000			
Name: V43, dtype: float64		Name: V43, dtype: float64		
Name: V43, dtype: float64				

The unique values in V43 feature is: 10

V43 feature has 168969 null values, which holds 0.2861262573238053 %

```
*****
*****
```

Analysis of V44 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	1.083891	mean	1.052407	
mean	1.963548			
std	0.639143	std	0.264426	
std	3.011257			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000			
50%	1.000000			
75%	1.000000			
75%	2.000000			
max	48.000000			
max	48.000000			
Name: V44, dtype: float64		Name: V44, dtype: float64		
Name: V44, dtype: float64				

The unique values in V44 feature is: 50

V44 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V45 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	421571.000000	count	407004.000000
count	14567.000000		
mean	1.120779	mean	1.081869
mean	2.207936		
std	0.729774	std	0.346760
std	3.290828		
min	0.000000	min	0.000000
min	0.000000		
25%	1.000000	25%	1.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	1.000000		
75%	1.000000	75%	1.000000
75%	2.000000		
max	48.000000	max	20.000000
max	48.000000		
Name: V45, dtype: float64		Name: V45, dtype: float64	
Name: V45, dtype: float64			

The unique values in V45 feature is: 50

V45 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V46 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	1.022286	mean	1.018312	
mean	1.133315			
std	0.166719	std	0.145434	
std	0.447973			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	6.000000	max	4.000000	
max	6.000000			
Name: V46, dtype: float64		Name: V46, dtype: float64		
Name: V46, dtype: float64				

The unique values in V46 feature is: 8

V46 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V47 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	1.038515	mean	1.031317	
mean	1.239651			
std	0.231862	std	0.190426	
std	0.707659			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	1.000000		50%	1.000000	
50%		1.000000				
	75%	1.000000		75%	1.000000	
75%		1.000000				
	max	12.000000		max	6.000000	
max		12.000000				
	Name: V47, dtype: float64			Name: V47, dtype: float64		
Name: V47, dtype: float64						
<hr/>						
<hr/>						

The unique values in V47 feature is: 10

V47 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V48 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.383174	mean	0.394306	
mean	0.072149			
std	0.508189	std	0.510969	
std	0.280881			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	0.000000			
max	5.000000	max	5.000000	
max	4.000000			
Name: V48, dtype: float64		Name: V48, dtype: float64		
Name: V48, dtype: float64				
<hr/>				
<hr/>				

The unique values in V48 feature is: 7

V48 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V49 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.397724	mean	0.409111	
mean	0.079563			
std	0.542654	std	0.545579	
std	0.317332			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	0.000000			
max	5.000000	max	5.000000	
max	5.000000			
Name: V49, dtype: float64		Name: V49, dtype: float64		
Name: V49, dtype: float64				

The unique values in V49 feature is: 7

V49 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V50 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.164746	mean	0.152013	
mean	0.520492			
std	0.373995	std	0.362182	
std	0.501791			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	0.000000	75%	0.000000
75%	1.000000		
max	5.000000	max	5.000000
max	4.000000		
Name: V50, dtype: float64		Name: V50, dtype: float64	
Name: V50, dtype: float64			

The unique values in V50 feature is: 7

V50 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V51 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.170579	mean	0.153524	
mean	0.647079			
std	0.403899	std	0.376541	
std	0.724260			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	1.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	6.000000	max	4.000000	
max	6.000000			
Name: V51, dtype: float64		Name: V51, dtype: float64		
Name: V51, dtype: float64				

The unique values in V51 feature is: 8

V51 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V52 feature

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	421571.000000	count	407004.000000	
count	14567.000000			
mean	0.182695	mean	0.162807	
mean	0.738381			
std	0.439002	std	0.396980	
std	0.924367			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	1.000000	50%	0.000000	
50%	1.000000			
75%	1.000000	75%	0.000000	
75%	1.000000			
max	12.000000	max	6.000000	
max	12.000000			
Name: V52, dtype: float64		Name: V52, dtype: float64		
Name: V52, dtype: float64				

The unique values in V52 feature is: 10

V52 feature has 168969 null values, which holds 0.2861262573238053 %

Analysis of V53 feature***Five Number Summary***

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.577586	mean	0.581956	
mean	0.443048			
std	0.511571	std	0.509111	
std	0.565967			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	1.000000	50%	1.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	5.000000	max	5.000000
max	3.000000		
Name: V53, dtype: float64		Name: V53, dtype: float64	
Name: V53, dtype: float64			

The unique values in V53 feature is: 7

V53 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V54 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.619982	mean	0.623550	
mean	0.510152			
std	0.534635	std	0.529922	
std	0.654092			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	1.000000	50%	1.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	6.000000	max	5.000000	
max	6.000000			
Name: V54, dtype: float64		Name: V54, dtype: float64		
Name: V54, dtype: float64				

The unique values in V54 feature is: 8

V54 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V55 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	513444.000000	497290.000000	
count	16154.000000		
mean	1.067670	1.064136	
mean	1.176489		
std	0.391364	0.382941	
std	0.584570		
min	0.000000	0.000000	
min	0.000000		
25%	1.000000	1.000000	
25%	1.000000		
50%	1.000000	1.000000	
50%	1.000000		
75%	1.000000	1.000000	
75%	1.000000		
max	17.000000	17.000000	
max	8.000000		
	Name: V55, dtype: float64	Name: V55, dtype: float64	
	Name: V55, dtype: float64		

The unique values in V55 feature is: 19

V55 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V56 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	513444.000000	497290.000000	
count	16154.000000		
mean	1.120979	1.113668	
mean	1.346044		
std	0.661129	0.634752	
std	1.198795		
min	0.000000	0.000000	
min	0.000000		
25%	1.000000	1.000000	
25%	1.000000		
50%	1.000000	1.000000	
50%	1.000000		
75%	1.000000	1.000000	
75%	1.000000		
max	51.000000	51.000000	
max	30.000000		

Name: V56, dtype: float64	Name: V56, dtype: float64
Name: V56, dtype: float64	
+-----+-----+	+-----+-----+
-----+-----+	

The unique values in V56 feature is: 53

V56 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V57 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.128312	mean	0.117686	
mean	0.455429			
std	0.349094	std	0.333895	
std	0.575313			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	6.000000	max	6.000000	
max	5.000000			
	Name: V57, dtype: float64	Name: V57, dtype: float64		
	Name: V57, dtype: float64			
+-----+-----+	+-----+-----+	+-----+-----+		
-----+-----+				

The unique values in V57 feature is: 8

V57 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V58 feature

Five Number Summary

+-----+-----+	+-----+-----+	+-----+-----+
-----+-----+		

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.132453	mean	0.120821	
mean	0.490529			
std	0.372907	std	0.354333	
std	0.650019			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	10.000000	max	10.000000	
max	7.000000			
Name: V58, dtype: float64		Name: V58, dtype: float64		
Name: V58, dtype: float64				

The unique values in V58 feature is: 12

V58 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V59 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.134433	mean	0.124107	
mean	0.452334			
std	0.379291	std	0.365292	
std	0.600346			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	16.000000	max	16.000000	
max	5.000000			
Name: V59, dtype: float64		Name: V59, dtype: float64		
Name: V59, dtype: float64				

-----+-----+
-----+-----+

The unique values in V59 feature is: 18
V59 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V60 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.142537	mean	0.131159	
mean	0.492819			
std	0.418230	std	0.401378	
std	0.688078			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	16.000000	max	16.000000	
max	9.000000			
Name: V60, dtype: float64		Name: V60, dtype: float64		
Name: V60, dtype: float64				

The unique values in V60 feature is: 18
V60 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V61 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			

count	513444.000000		count	497290.000000
count	16154.000000			
mean	0.829785		mean	0.825896
mean	0.949486			
std	0.436617		std	0.433733
std	0.503119			
min	0.000000		min	0.000000
min	0.000000			
25%	1.000000		25%	1.000000
25%	1.000000			
50%	1.000000		50%	1.000000
50%	1.000000			
75%	1.000000		75%	1.000000
75%	1.000000			
max	6.000000		max	6.000000
max	6.000000			
Name: V61, dtype: float64			Name: V61, dtype: float64	
Name: V61, dtype: float64				

The unique values in V61 feature is: 8

V61 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V62 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.867563	mean	0.861250	
mean	1.061904			
std	0.483803	std	0.474493	
std	0.685393			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	10.000000	max	10.000000	
max	7.000000			
Name: V62, dtype: float64		Name: V62, dtype: float64		
Name: V62, dtype: float64				

The unique values in V62 feature is: 12
 V62 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V63 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.130743	mean	0.120923	
mean	0.433020			
std	0.355262	std	0.342257	
std	0.557809			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	4.000000			
Name: V63, dtype: float64		Name: V63, dtype: float64		
Name: V63, dtype: float64				

The unique values in V63 feature is: 9
 V63 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V64 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.130743	mean	0.120923	
mean	0.433020			
std	0.355262	std	0.342257	
std	0.557809			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	4.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

count	513444.000000	count	497290.000000
count	16154.000000		
mean	0.141825	mean	0.130817
mean	0.480686		
std	0.406157	std	0.390062
std	0.664032		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	1.000000		
max	7.000000	max	7.000000
max	7.000000		
Name: V64, dtype: float64		Name: V64, dtype: float64	
Name: V64, dtype: float64			

The unique values in V64 feature is: 9

V64 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V65 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.999663	mean	0.999654	
mean	0.999938			
std	0.018353	std	0.018595	
std	0.007868			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	1.000000	max	1.000000	
max	1.000000			
Name: V65, dtype: float64		Name: V65, dtype: float64		
Name: V65, dtype: float64				

The unique values in V65 feature is: 3
 V65 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V66 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.981038	mean	0.979571	
mean	1.026185			
std	0.216235	std	0.209554	
std	0.363635			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	6.000000			
Name: V66, dtype: float64		Name: V66, dtype: float64		
Name: V66, dtype: float64				

The unique values in V66 feature is: 9
 V66 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V67 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	0.998121	mean	0.994744
mean	1.102080	std	0.233315
std	0.245912	std	0.233315
std	0.484941	min	0.000000
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
25%	1.000000	50%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
75%	1.000000	max	7.000000
max	8.000000	max	7.000000
	Name: V67, dtype: float64	Name: V67, dtype: float64	
	Name: V67, dtype: float64		

The unique values in V67 feature is: 10

V67 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V68 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	513444.000000	count	497290.000000
count	16154.0	mean	0.000551
mean	0.0	std	0.024142
std	0.0	min	0.000000
min	0.0	25%	0.000000
25%	0.0	50%	0.000000
50%	0.0	75%	0.000000
75%	0.0	max	2.000000
max	0.0		
	Name: V68, dtype: float64	Name: V68, dtype: float64	
	Name: V68, dtype: float64		

The unique values in V68 feature is: 4

V68 feature has 77096 null values, which holds 0.13055169844549055 %

```
*****
*****
```

Analysis of V69 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.390200	mean	0.399877	
mean	0.092299			
std	0.513696	std	0.515878	
std	0.321093			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	0.000000			
max	5.000000	max	4.000000	
max	5.000000			
Name: V69, dtype: float64		Name: V69, dtype: float64		
Name: V69, dtype: float64				

The unique values in V69 feature is: 7

V69 feature has 77096 null values, which holds 0.13055169844549055 %

```
*****
*****
```

Analysis of V70 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.407924	mean	0.418008	
mean	0.097499			

1. self_case_study_1_IEEE_fraud_detection_EDA

std	0.554499	std	0.557105
std	0.344715	min	0.000000
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	1.000000
75%	0.000000	75%	1.000000
max	6.000000	max	5.000000
max	6.000000	Name: V70, dtype: float64	Name: V70, dtype: float64
Name: V70, dtype: float64		Name: V70, dtype: float64	

The unique values in V70 feature is: 8

V70 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V71 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.140639	mean	0.130340	
mean	0.457658			
std	0.364303	std	0.350576	
std	0.575374			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	6.000000	max	6.000000	
max	5.000000			
Name: V71, dtype: float64		Name: V71, dtype: float64		
Name: V71, dtype: float64				

The unique values in V71 feature is: 8

V71 feature has 77096 null values, which holds 0.13055169844549055 %

```
*****
*****
```

Analysis of V72 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.145124	mean	0.133833	
mean	0.492695			
std	0.389212	std	0.372346	
std	0.649762			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	10.000000	max	10.000000	
max	7.000000			
Name: V72, dtype: float64		Name: V72, dtype: float64		
Name: V72, dtype: float64				

The unique values in V72 feature is: 12

V72 feature has 77096 null values, which holds 0.13055169844549055 %

```
*****
*****
```

Analysis of V73 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	513444.000000	count	497290.000000	
count	16154.000000			
mean	0.139982	mean	0.128478	
mean	0.494119			
std	0.367098	std	0.350753	
std	0.605386			

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000		min	0.000000	
min	0.000000				
25%	0.000000		25%	0.000000	
25%	0.000000				
50%	0.000000		50%	0.000000	
50%	0.000000				
75%	0.000000		75%	0.000000	
75%	1.000000				
max	7.000000		max	7.000000	
max	6.000000				
Name: V73, dtype: float64		Name: V73, dtype: float64			
Name: V73, dtype: float64					

The unique values in V73 feature is: 9

V73 feature has 77096 null values, which holds 0.13055169844549055 %

Analysis of V74 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature	
Count Description of fraud transactions					
count	513444.000000		count	497290.000000	
count	16154.000000				
mean	0.152147		mean	0.138969	
mean	0.557818				
std	0.393274		std	0.371890	
std	0.698917				
min	0.000000		min	0.000000	
min	0.000000				
25%	0.000000		25%	0.000000	
25%	0.000000				
50%	0.000000		50%	0.000000	
50%	0.000000				
75%	0.000000		75%	0.000000	
75%	1.000000				
max	8.000000		max	7.000000	
max	8.000000				
Name: V74, dtype: float64		Name: V74, dtype: float64			
Name: V74, dtype: float64					

The unique values in V74 feature is: 10

V74 feature has 77096 null values, which holds 0.13055169844549055 %

```
*****
*****
```

Analysis of V75 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.544278	mean	0.549210	
mean	0.398304			
std	0.514318	std	0.512495	
std	0.545796			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	1.000000	50%	1.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	4.000000	max	4.000000	
max	3.000000			
Name: V75, dtype: float64		Name: V75, dtype: float64		
Name: V75, dtype: float64				

The unique values in V75 feature is: 6

V75 feature has 89164 null values, which holds 0.15098723202492634 %

```
*****
*****
```

Analysis of V76 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.587557	mean	0.592150	
mean	0.451638			
std	0.538230	std	0.534486	
std	0.624103			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	0.000000	25%	0.000000	
25%	0.000000			
50%	1.000000	50%	1.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	6.000000	max	5.000000	
max	6.000000			
Name: V76, dtype: float64		Name: V76, dtype: float64		
Name: V76, dtype: float64				

The unique values in V76 feature is: 8

V76 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V77 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	1.086893	mean	1.071377	
mean	1.546104			
std	0.532958	std	0.437279	
std	1.677384			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	30.000000	max	19.000000	
max	30.000000			
Name: V77, dtype: float64		Name: V77, dtype: float64		
Name: V77, dtype: float64				

The unique values in V77 feature is: 32

V77 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V78 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	1.144462	mean	1.122854	
mean	1.783975			
std	0.781474	std	0.654110	
std	2.366350			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			
max	31.000000	max	25.000000	
max	31.000000			
Name: V78, dtype: float64		Name: V78, dtype: float64		
Name: V78, dtype: float64				

The unique values in V78 feature is: 33

V78 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V79 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.136867	mean	0.123310	
mean	0.538109			
std	0.379904	std	0.358534	
std	0.667046			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	0.000000		50%	0.000000	
50%		0.000000				
	75%	0.000000		75%	0.000000	
75%		1.000000				
	max	7.000000		max	7.000000	
max		6.000000				
	Name: V79, dtype: float64			Name: V79, dtype: float64		
Name: V79, dtype: float64						
<hr/>						
<hr/>						

The unique values in V79 feature is: 9

V79 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V80 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.143954	mean	0.130368	
mean	0.546043			
std	0.409932	std	0.377840	
std	0.865550			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	19.000000	max	15.000000	
max	19.000000			
	Name: V80, dtype: float64		Name: V80, dtype: float64	
Name: V80, dtype: float64				
<hr/>				
<hr/>				

The unique values in V80 feature is: 21

V80 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V81 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.152696	mean	0.137533	
mean	0.601452			
std	0.452298	std	0.415043	
std	0.976099			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	19.000000	max	15.000000	
max	19.000000			
Name: V81, dtype: float64		Name: V81, dtype: float64		
Name: V81, dtype: float64				

The unique values in V81 feature is: 21

V81 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V82 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.844610	mean	0.840640	
mean	0.962104			
std	0.422529	std	0.419453	
std	0.490821			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	1.000000	75%	1.000000
75%	1.000000		
max	7.000000	max	7.000000
max	6.000000		
Name: V82, dtype: float64		Name: V82, dtype: float64	
Name: V82, dtype: float64			

The unique values in V82 feature is: 9

V82 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V83 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.881965	mean	0.875981	
mean	1.059071			
std	0.470757	std	0.463645	
std	0.621182			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	6.000000			
Name: V83, dtype: float64		Name: V83, dtype: float64		
Name: V83, dtype: float64				

The unique values in V83 feature is: 9

V83 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V84 feature

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.137145	mean	0.125896	
mean	0.470068			
std	0.361875	std	0.348039	
std	0.554162			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	4.000000			
Name: V84, dtype: float64		Name: V84, dtype: float64		
Name: V84, dtype: float64				

The unique values in V84 feature is: 9

V84 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V85 feature***Five Number Summary***

Feature Description of fraud transactions		Feature Description of legit transactions		Feature Description
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.149788	mean	0.136770	
mean	0.535058			
std	0.419445	std	0.399721	
std	0.707608			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	7.000000	max	7.000000
max	7.000000		
Name: V85, dtype: float64		Name: V85, dtype: float64	
Name: V85, dtype: float64			

The unique values in V85 feature is: 9

V85 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V86 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	1.064885	mean	1.045453	
mean	1.640020			
std	0.419807	std	0.261994	
std	1.737494			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	30.000000	max	15.000000	
max	30.000000			
Name: V86, dtype: float64		Name: V86, dtype: float64		
Name: V86, dtype: float64				

The unique values in V86 feature is: 32

V86 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V87 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	501376.000000	count	484989.000000
count	16387.000000		
mean	1.099456	mean	1.075798
mean	1.799658		
std	0.511281	std	0.339411
std	2.020390		
min	0.000000	min	0.000000
min	0.000000		
25%	1.000000	25%	1.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	1.000000		
75%	1.000000	75%	1.000000
75%	2.000000		
max	30.000000	max	19.000000
max	30.000000		
Name: V87, dtype: float64		Name: V87, dtype: float64	
Name: V87, dtype: float64			

The unique values in V87 feature is: 32

V87 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V88 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	501376.000000	count	484989.000000
count	16387.000000		
mean	0.999246	mean	0.999243
mean	0.999329		
std	0.027447	std	0.027498
std	0.025901		
min	0.000000	min	0.000000
min	0.000000		
25%	1.000000	25%	1.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	1.000000		
75%	1.000000	75%	1.000000
75%	1.000000		
max	1.000000	max	1.000000
max	1.000000		

Name: V88, dtype: float64	Name: V88, dtype: float64
Name: V88, dtype: float64	
+-----+-----+	+-----+-----+
-----+-----+	

The unique values in V88 feature is: 3

V88 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V89 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.0			
mean	0.000902	mean	0.000932	
mean	0.0			
std	0.031943	std	0.032478	
std	0.0			
min	0.000000	min	0.000000	
min	0.0			
25%	0.000000	25%	0.000000	
25%	0.0			
50%	0.000000	50%	0.000000	
50%	0.0			
75%	0.000000	75%	0.000000	
75%	0.0			
max	2.000000	max	2.000000	
max	0.0			
	Name: V89, dtype: float64		Name: V89, dtype: float64	
	Name: V89, dtype: float64			
+-----+-----+	+-----+-----+	+-----+-----+	+-----+-----+	
-----+-----+				

The unique values in V89 feature is: 4

V89 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V90 feature

Five Number Summary

+-----+-----+	+-----+-----+	+-----+-----+
-----+-----+		

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	501376.000000	count	484989.000000
count	16387.000000		
mean	0.401862	mean	0.412315
mean	0.092512		
std	0.516153	std	0.518220
std	0.322831		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	1.000000	75%	1.000000
75%	0.000000		
max	5.000000	max	5.000000
max	5.000000		
Name: V90, dtype: float64		Name: V90, dtype: float64	
Name: V90, dtype: float64			

The unique values in V90 feature is: 7

V90 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V91 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	501376.000000	count	484989.000000
count	16387.000000		
mean	0.420461	mean	0.431352
mean	0.098127		
std	0.561865	std	0.564455
std	0.349208		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	1.000000	75%	1.000000
75%	0.000000		
max	6.000000	max	6.000000
max	6.000000		
Name: V91, dtype: float64		Name: V91, dtype: float64	
Name: V91, dtype: float64			

```
+-----+-----+
-----+
```

The unique values in V91 feature is: 8
V91 feature has 89164 null values, which holds 0.15098723202492634 %

```
*****  
*****
```

Analysis of V92 feature

Five Number Summary

```
+-----+-----+
-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+
-----+
| count      501376.000000 | | count      484989.000000 | |
count      16387.000000 | | mean       0.137857 | |
| mean       0.150322 | | mean       0.519253 | |
mean       0.375682 | | std        0.358622 | |
std        0.609271 | | min        0.000000 | |
| min        0.000000 | | max        7.000000 | |
min        0.000000 | | 25%        0.000000 | |
| 25%        0.000000 | | 50%        0.000000 | |
25%        0.000000 | | 75%        0.000000 | |
| 50%        0.000000 | | max        6.000000 | |
50%        0.000000 | | 75%        1.000000 | |
| 75%        0.000000 | | max        7.000000 | |
75%        1.000000 | | 25%        0.000000 | |
| max        7.000000 | | 50%        0.000000 | |
max        6.000000 | | 75%        0.000000 | |
| Name: V92, dtype: float64 | | Name: V92, dtype: float64 | |
Name: V92, dtype: float64 | +-----+
+-----+-----+
-----+
```

The unique values in V92 feature is: 9
V92 feature has 89164 null values, which holds 0.15098723202492634 %

```
*****  
*****
```

Analysis of V93 feature

Five Number Summary

```
+-----+-----+
-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+
-----+
```

	count	501376.000000	count	484989.000000
	count	16387.000000		
	mean	0.154812	mean	0.141428
	mean	0.550925		
	std	0.402314	std	0.382908
	std	0.671300		
	min	0.000000	min	0.000000
	min	0.000000		
	25%	0.000000	25%	0.000000
	25%	0.000000		
	50%	0.000000	50%	0.000000
	50%	0.000000		
	75%	0.000000	75%	0.000000
	75%	1.000000		
	max	7.000000	max	7.000000
	max	6.000000		
Name: V93, dtype: float64		Name: V93, dtype: float64		
Name: V93, dtype: float64				

The unique values in V93 feature is: 9

V93 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V94 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	501376.000000	count	484989.000000	
count	16387.000000			
mean	0.137007	mean	0.125098	
mean	0.489473			
std	0.343878	std	0.330855	
std	0.499904			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	2.000000	max	2.000000	
max	1.000000			
Name: V94, dtype: float64		Name: V94, dtype: float64		
Name: V94, dtype: float64				

The unique values in V94 feature is: 4
V94 feature has 89164 null values, which holds 0.15098723202492634 %

Analysis of V95 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.038019	mean	1.054517	
mean	0.582873			
std	21.034304	std	21.410652	
std	1.239608			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	880.000000	max	880.000000	
max	13.000000			
Name: V95, dtype: float64		Name: V95, dtype: float64		
Name: V95, dtype: float64				

The unique values in V95 feature is: 882
V95 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V96 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.038019	mean	1.054517	
mean	0.582873			
std	21.034304	std	21.410652	
std	1.239608			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	880.000000	max	880.000000	
max	13.000000			
Name: V96, dtype: float64		Name: V96, dtype: float64		
Name: V96, dtype: float64				

1. self_case_study_1_IEEE_fraud_detection_EDA

count 590226.00000	count 569580.00000	
count 20646.00000		
mean 3.00519	mean 3.047207	
mean 1.846023		
std 40.23949	std 40.952917	
std 4.449918		
min 0.00000	min 0.000000	
min 0.000000		
25% 0.00000	25% 0.000000	
25% 0.000000		
50% 0.00000	50% 0.000000	
50% 0.000000		
75% 1.00000	75% 1.000000	
75% 2.000000		
max 1410.00000	max 1410.000000	
max 55.000000		
Name: V96, dtype: float64 Name: V96, dtype: float64		
Name: V96, dtype: float64		
+-----+-----+-----+		
- - - - +		

The unique values in V96 feature is: 1411

V96 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V97 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count 590226.00000		count 569580.00000		
count 20646.00000				
mean 1.718933		mean 1.742616		
mean 1.065582				
std 27.700449		std 28.194494		
std 2.245991				
min 0.000000		min 0.000000		
min 0.000000				
25% 0.000000		25% 0.000000		
25% 0.000000				
50% 0.000000		50% 0.000000		
50% 0.000000				
75% 0.000000		75% 0.000000		
75% 1.000000				
max 976.000000		max 976.000000		
max 26.000000				
Name: V97, dtype: float64 Name: V97, dtype: float64				
Name: V97, dtype: float64				
+-----+-----+-----+				
- - - - +				

The unique values in V97 feature is: 977
 V97 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V98 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	0.061985	mean	0.060502	
mean	0.102877			
std	0.284995	std	0.281261	
std	0.371254			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	12.000000	max	12.000000	
max	6.000000			
Name: V98, dtype: float64		Name: V98, dtype: float64		
Name: V98, dtype: float64				

The unique values in V98 feature is: 14
 V98 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V99 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	0.894986	mean	0.901817
mean	0.706529	std	2.746030
std	2.722552	std	1.957960
min	0.000000	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	1.000000	75%	1.000000
75%	0.000000	max	88.000000
max	88.000000	max	24.000000
Name: V99, dtype: float64		Name: V99, dtype: float64	
Name: V99, dtype: float64			

The unique values in V99 feature is: 90

V99 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V100 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	0.273504	mean	0.271468	
mean	0.329652	std	0.947877	
std	0.947176	std	0.925890	
min	0.000000	min	0.000000	
min	0.000000	25%	0.000000	
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
50%	0.000000	75%	0.000000	
75%	0.000000	75%	0.000000	
max	28.000000	max	28.000000	
max	17.000000			
Name: V100, dtype: float64		Name: V100, dtype: float64		
Name: V100, dtype: float64				

The unique values in V100 feature is: 30

V100 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V101 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	590226.000000	count	569580.000000
mean	0.889249	mean	0.907714
std	20.582571	std	20.951164
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	869.000000	max	869.000000
Name: V101, dtype: float64		Name: V101, dtype: float64	
Name: V101, dtype: float64			

The unique values in V101 feature is: 871

V101 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V102 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	590226.000000	count	569580.000000
mean	1.827229	mean	1.860388
mean	0.912429		

1. self_case_study_1_IEEE_fraud_detection_EDA

std	35.928456	std	36.569617
std	2.760635	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	1.000000	max	1285.000000
max	41.000000	Name: V102, dtype: float64	Name: V102, dtype: float64
Name: V102, dtype: float64			

The unique values in V102 feature is: 1286

V102 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V103 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.279288	mean	1.30433	
mean	0.588443			
std	25.685642	std	26.14497	
std	1.570681			
min	0.000000	min	0.00000	
min	0.000000			
25%	0.000000	25%	0.00000	
25%	0.000000			
50%	0.000000	50%	0.00000	
50%	0.000000			
75%	0.000000	75%	0.00000	
75%	0.000000			
max	928.000000	max	928.00000	
max	21.000000			
Name: V103, dtype: float64		Name: V103, dtype: float64		
Name: V103, dtype: float64				

The unique values in V103 feature is: 929

V103 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V104 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590226.000000	569580.000000	count
mean	0.085433	0.085003	mean
std	0.097307	0.653947	std
min	0.648718	0.482461	min
25%	0.000000	0.000000	25%
50%	0.000000	0.000000	50%
75%	0.000000	0.000000	75%
max	15.000000	15.000000	max
	12.000000		
Name: V104, dtype: float64		Name: V104, dtype: float64	
Name: V104, dtype: float64			

The unique values in V104 feature is: 17

V104 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V105 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590226.000000	569580.000000	count
mean	0.281145	0.283219	mean
std	0.223917	3.426548	std
min	3.373948	1.229567	min
25%	0.000000	0.000000	25%
50%	0.000000	0.000000	50%
75%	0.000000	0.000000	75%
max	15.000000	15.000000	max
	12.000000		
Name: V105, dtype: float64		Name: V105, dtype: float64	
Name: V105, dtype: float64			

1. self_case_study_1_IEEE_fraud_detection_EDA

	min	0.000000		min	0.000000	
min		0.000000				
	25%	0.000000		25%	0.000000	
25%		0.000000				
	50%	0.000000		50%	0.000000	
50%		0.000000				
	75%	0.000000		75%	0.000000	
75%		0.000000				
	max	99.000000		max	99.000000	
max		27.000000				
	Name: V105, dtype: float64			Name: V105, dtype: float64		
	Name: V105, dtype: float64					+-----+
			-+-----+			

The unique values in V105 feature is: 101

V105 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V106 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000		count	569580.000000
count	20646.000000			
mean	0.164584		mean	0.165311
mean	0.144532			
std	1.825665		std	1.852491
std	0.781436			
min	0.000000		min	0.000000
min	0.000000			
25%	0.000000		25%	0.000000
25%	0.000000			
50%	0.000000		50%	0.000000
50%	0.000000			
75%	0.000000		75%	0.000000
75%	0.000000			
max	55.000000		max	55.000000
max	23.000000			
	Name: V106, dtype: float64		Name: V106, dtype: float64	
	Name: V106, dtype: float64			+-----+
			-+-----+	

The unique values in V106 feature is: 57

V106 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V107 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	0.999580	mean	0.999570	
mean	0.999855			
std	0.020494	std	0.020735	
std	0.012054			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	1.000000	max	1.000000	
max	1.000000			
Name: V107, dtype: float64		Name: V107, dtype: float64		
Name: V107, dtype: float64				

The unique values in V107 feature is: 3

V107 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V108 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.004613	mean	1.003118	
mean	1.045868			
std	0.081146	std	0.064075	
std	0.270586			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	1.000000		25%	1.000000	
25%	1.000000		50%	1.000000	
50%	1.000000		50%	1.000000	
75%	1.000000		75%	1.000000	
75%	1.000000		max	4.000000	
max	7.000000		max	4.000000	
Name: V108, dtype: float64		Name: V108, dtype: float64			
Name: V108, dtype: float64					

The unique values in V108 feature is: 9

V108 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V109 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000		count	569580.000000
count	20646.000000		mean	1.013333
mean	1.014816		std	0.117687
mean	1.055749		min	0.000000
std	0.127769		min	0.000000
std	0.287870		25%	1.000000
min	0.000000		25%	1.000000
min	1.000000		50%	1.000000
25%	1.000000		50%	1.000000
50%	1.000000		75%	1.000000
75%	1.000000		75%	1.000000
max	7.000000		max	4.000000
max	7.000000		max	4.000000
Name: V109, dtype: float64		Name: V109, dtype: float64		
Name: V109, dtype: float64				

The unique values in V109 feature is: 9

V109 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V110 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.007739	mean	1.006227	
mean	1.049453			
std	0.097290	std	0.083539	
std	0.276163			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	7.000000	max	4.000000	
max	7.000000			
Name: V110, dtype: float64		Name: V110, dtype: float64		
Name: V110, dtype: float64				

The unique values in V110 feature is: 9

V110 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V111 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.002563	mean	1.000923	
mean	1.047806			
std	0.070830	std	0.044016	
std	0.296407			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	1.000000		50%	1.000000	
50%		1.000000				
	75%	1.000000		75%	1.000000	
75%		1.000000				
	max	9.000000		max	4.000000	
max		9.000000				
	Name: V111, dtype: float64			Name: V111, dtype: float64		
Name: V111, dtype: float64						
+-----+-----+-----+-----+-----+-----+						

The unique values in V111 feature is: 11

V111 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V112 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	590226.000000		count	569580.000000	
count	20646.000000				
mean	1.005356		mean	1.003661	
mean	1.052117				
std	0.084617		std	0.063399	
std	0.302562				
min	0.000000		min	0.000000	
min	1.000000				
25%	1.000000		25%	1.000000	
25%	1.000000				
50%	1.000000		50%	1.000000	
50%	1.000000				
75%	1.000000		75%	1.000000	
75%	1.000000				
max	9.000000		max	4.000000	
max	9.000000				
	Name: V112, dtype: float64			Name: V112, dtype: float64	
Name: V112, dtype: float64					
+-----+-----+-----+-----+-----+-----+					

The unique values in V112 feature is: 11

V112 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V113 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.003383	mean	1.001733	
mean	1.048920			
std	0.074754	std	0.049953	
std	0.297938			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	9.000000	max	4.000000	
max	9.000000			
Name: V113, dtype: float64		Name: V113, dtype: float64		
Name: V113, dtype: float64				

The unique values in V113 feature is: 11

V113 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V114 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.009298	mean	1.007283	
mean	1.064904			
std	0.110179	std	0.092617	
std	0.327406			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	1.000000	75%	1.000000
75%	1.000000		
max	6.000000	max	5.000000
max	6.000000		
Name: V114, dtype: float64		Name: V114, dtype: float64	
Name: V114, dtype: float64			

The unique values in V114 feature is: 8
V114 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V115 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	569580.00000	
count	20646.00000	
mean	1.030380	
mean	1.089557	
std	0.180506	
std	0.365993	
min	0.000000	
min	1.000000	
25%	1.000000	
25%	1.000000	
50%	1.000000	
50%	1.000000	
75%	1.000000	
75%	1.000000	
max	5.000000	
max	6.000000	
Name: V115, dtype: float64		Name: V115, dtype: float64
Name: V115, dtype: float64		

The unique values in V115 feature is: 8
V115 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V116 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.015738	mean	1.013638	
mean	1.07367			
std	0.136770	std	0.122735	
std	0.34017			
min	0.000000	min	0.000000	
min	0.00000			
25%	1.000000	25%	1.000000	
25%	1.00000			
50%	1.000000	50%	1.000000	
50%	1.00000			
75%	1.000000	75%	1.000000	
75%	1.00000			
max	6.000000	max	5.000000	
max	6.00000			
Name: V116, dtype: float64		Name: V116, dtype: float64		
Name: V116, dtype: float64				

The unique values in V116 feature is: 8

V116 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V117 feature***Five Number Summary***

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.000391	mean	1.000060	
mean	1.009542			
std	0.035238	std	0.030504	
std	0.098701			
min	0.000000	min	0.000000	
min	0.00000			
25%	1.000000	25%	1.000000	
25%	1.00000			
50%	1.000000	50%	1.000000	
50%	1.00000			
75%	1.000000	75%	1.000000	
75%	1.00000			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	3.000000	max	3.000000
max	2.000000		
Name: V117, dtype: float64		Name: V117, dtype: float64	
Name: V117, dtype: float64			

The unique values in V117 feature is: 5
V117 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V118 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	1.001474	mean	1.001157	
mean	1.010220			
std	0.041011	std	0.037059	
std	0.100578			
min	0.000000	min	0.000000	
min	1.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	3.000000	max	3.000000	
max	2.000000			
Name: V118, dtype: float64		Name: V118, dtype: float64		
Name: V118, dtype: float64				

The unique values in V118 feature is: 5
V118 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V119 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	590226.000000	count	569580.000000
count	20646.000000	mean	1.000395
mean	1.000729	std	0.031770
std	1.009929	min	0.000000
std	0.036392	25%	1.000000
min	0.099640	50%	1.000000
min	0.000000	75%	1.000000
25%	0.000000	max	3.000000
50%	1.000000	max	2.000000
75%	1.000000	Name: V119, dtype: float64	Name: V119, dtype: float64
max	3.000000	Name: V119, dtype: float64	
max	2.000000		

The unique values in V119 feature is: 5

V119 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V120 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	590226.000000	count	569580.000000
count	20646.000000	mean	1.000718
mean	1.000874	std	0.039766
std	1.005183	min	0.000000
std	0.041684	25%	1.000000
min	0.077639	50%	1.000000
min	0.000000	75%	1.000000
25%	0.000000	max	2.000000
50%	1.000000		
50%	1.000000		
75%	1.000000		
75%	1.000000		
max	3.000000		
max	3.000000		

```
| Name: V120, dtype: float64 |           Name: V120, dtype: float64 |
| Name: V120, dtype: float64 |           |
+-----+-----+-----+
```

The unique values in V120 feature is: 5
V120 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V121 feature

Five Number Summary

```
+-----+-----+-----+
|-----+-----+-----+
| Feature Description      | Feature Description of legit transactions | Feature
| Description of fraud transactions |                               |
+-----+-----+-----+
-----+
| count      590226.000000 |           count      569580.000000 |
count      20646.000000 |           |           mean      1.004119 |
| mean       1.004276 |           |           std       0.065830 |
mean       1.008622 |           |           min       0.000000 |
| std        0.067097 |           |           |           0.000000 |
std        0.095545 |           |           |           1.000000 |
| min        0.000000 |           |           |           1.000000 |
min        1.000000 |           |           |           1.000000 |
| 25%        1.000000 |           |           25%       1.000000 |
25%        1.000000 |           |           |           1.000000 |
| 50%        1.000000 |           |           50%       1.000000 |
50%        1.000000 |           |           |           1.000000 |
| 75%        1.000000 |           |           75%       1.000000 |
75%        1.000000 |           |           |           1.000000 |
| max        3.000000 |           |           max       3.000000 |
max        3.000000 |           |           |           3.000000 |
| Name: V121, dtype: float64 |           Name: V121, dtype: float64 |
Name: V121, dtype: float64 |           |
+-----+-----+-----+
-----+
```

The unique values in V121 feature is: 5
V121 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V122 feature

Five Number Summary

```
+-----+-----+-----+
-----+
```

1. self_case_study_1_IEEE_fraud_detection_EDA

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	590226.000000	count	569580.000000
mean	1.001759	mean	1.001585
std	0.048636	std	0.046801
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	3.000000	max	2.000000
	Name: V122, dtype: float64	Name: V122, dtype: float64	
	Name: V122, dtype: float64		

The unique values in V122 feature is: 5

V122 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V123 feature

Five Number Summary

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	590226.000000	count	569580.000000
mean	1.031120	mean	1.025117
std	0.228134	std	0.177467
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	13.000000	max	9.000000
	Name: V123, dtype: float64	Name: V123, dtype: float64	
	Name: V123, dtype: float64		

```
+-----+-----+
-----+-----+
```

The unique values in V123 feature is: 15
V123 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****  
*****
```

Analysis of V124 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description      | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+-----+
| count      590226.000000 |           count      569580.000000 |
count      20646.000000 |           mean       1.087510 |
mean       1.243728 |           std        0.347368 |
std        0.805333 |           min       0.000000 |
min        1.000000 |           25%       1.000000 |
25%        1.000000 |           50%       1.000000 |
50%        1.000000 |           75%       1.000000 |
75%        1.000000 |           max       11.000000 |
max        13.000000 |           Name: V124, dtype: float64 |
Name: V124, dtype: float64 |           Name: V124, dtype: float64 |
+-----+-----+-----+
```

The unique values in V124 feature is: 15
V124 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****  
*****
```

Analysis of V125 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description      | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+-----+
```

count	590226.000000	count	569580.000000
count	20646.000000	mean	1.044443
mean	1.050415	std	0.240997
mean	1.215151	min	0.000000
std	0.280037	25%	1.000000
std	0.781981	50%	1.000000
min	0.000000	75%	1.000000
min	0.000000	max	9.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	13.000000	max	9.000000
	Name: V125, dtype: float64		Name: V125, dtype: float64
	Name: V125, dtype: float64		

The unique values in V125 feature is: 15

V125 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V126 feature

Five Number Summary

		Feature Description of legit transactions	Feature Description of fraud transactions
count	590226.000000	count	569580.000000
count	20646.000000	mean	131.563693
mean	129.979417	std	2388.479348
std	86.272554	min	284.631132
std	2346.951681	25%	0.000000
min	0.000000	50%	0.000000
25%	0.000000	75%	0.000000
50%	0.000000	max	40.000000
75%	0.000000	max	160000.000000
max	6625.000000		
	Name: V126, dtype: float64		Name: V126, dtype: float64
	Name: V126, dtype: float64		

The unique values in V126 feature is: 10300
 V126 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V127 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590226.000000	569580.000000	
mean	336.611559	338.732854	
std	4238.666949	4311.400223	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	107.949997	107.949997	
max	162.579044	160000.000000	
		160000.000000	
Name: V127, dtype: float64		Name: V127, dtype: float64	
Name: V127, dtype: float64			

The unique values in V127 feature is: 24415
 V127 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V128 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590226.000000	569580.000000	
mean	336.611559	338.732854	
std	4238.666949	4311.400223	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	107.949997	107.949997	
max	162.579044	160000.000000	
		160000.000000	
Name: V128, dtype: float64		Name: V128, dtype: float64	
Name: V128, dtype: float64			

1. self_case_study_1_IEEE_fraud_detection_EDA

count 590226.000000	count 569580.000000
count 20646.000000	
mean 204.094038	mean 205.200301
mean 173.574554	
std 3010.258774	std 3062.462466
std 561.111545	
min 0.000000	min 0.000000
min 0.000000	
25% 0.000000	25% 0.000000
25% 0.000000	
50% 0.000000	50% 0.000000
50% 0.000000	
75% 100.000000	75% 0.000000
75% 100.000000	
max 160000.000000	max 160000.000000
max 10162.000000	
Name: V128, dtype: float64 Name: V128, dtype: float64	
Name: V128, dtype: float64	

The unique values in V128 feature is: 14508

V128 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V129 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count 590226.000000		count 569580.000000		
count 20646.000000				
mean 8.768944	mean 8.493041			
mean 16.380528				
std 113.832828	std 114.773011			
std 83.478865				
min 0.000000	min 0.000000			
min 0.000000				
25% 0.000000	25% 0.000000			
25% 0.000000				
50% 0.000000	50% 0.000000			
50% 0.000000				
75% 0.000000	75% 0.000000			
75% 0.000000				
max 55125.000000	max 55125.000000			
max 1768.000000				
Name: V129, dtype: float64 Name: V129, dtype: float64				
Name: V129, dtype: float64				

The unique values in V129 feature is: 1969
 V129 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V130 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	92.165849	mean	91.980813	
mean	97.270605			
std	315.960485	std	315.485415	
std	328.762971			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	59.000000	75%	59.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	4350.000000			
Name: V130, dtype: float64		Name: V130, dtype: float64		
Name: V130, dtype: float64				

The unique values in V130 feature is: 12333
 V130 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V131 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	31.133302	mean	30.366999
mean	52.274012	std	160.045517
std	161.161258	min	0.000000
std	188.145066	25%	0.000000
min	0.000000	50%	0.000000
25%	0.000000	75%	0.000000
50%	0.000000	max	55125.000000
75%	0.000000		
max	55125.000000	Name: V131, dtype: float64	Name: V131, dtype: float64
	3981.000000	Name: V131, dtype: float64	
		+-----+	+-----+
		- - - - -	- - - - -

The unique values in V131 feature is: 4445

V131 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V132 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	103.513188	mean	105.352467	
mean	52.771340	std	2306.414902	
std	2266.106140	std	218.583926	
min	0.000000	min	0.000000	
min	0.000000	25%	0.000000	
25%	0.000000	50%	0.000000	
50%	0.000000	75%	0.000000	
75%	0.000000	max	93736.000000	
max	93736.000000	max	93736.000000	
	6098.000000			
		Name: V132, dtype: float64	Name: V132, dtype: float64	
		Name: V132, dtype: float64		
		+-----+	+-----+	+-----+
		- - - - -	- - - - -	- - - - -

The unique values in V132 feature is: 6561

V132 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V133 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	590226.000000	count	569580.000000
mean	204.889160	mean	207.176966
std	3796.316755	std	3862.820432
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	23.422851	max	133915.000000
count	15607.000000	count	133915.000000
Name: V133, dtype: float64		Name: V133, dtype: float64	
Name: V133, dtype: float64			

The unique values in V133 feature is: 9950

V133 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V134 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	590226.000000	count	569580.000000
mean	145.972328	mean	147.826416
std	94.821907	std	96.744160
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	23.422851	max	133915.000000
count	15607.000000	count	133915.000000
Name: V134, dtype: float64		Name: V134, dtype: float64	
Name: V134, dtype: float64			

1. self_case_study_1_IEEE_fraud_detection_EDA

std	2772.986817	std	2821.743297
std	401.688406	min	0.000000
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	0.000000	max	98476.000000
max	98476.000000	max	98476.000000
	Name: V134, dtype: float64		Name: V134, dtype: float64
	Name: V134, dtype: float64		

The unique values in V134 feature is: 8179

V134 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V135 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	590226.000000	count	569580.000000
count	20646.000000	mean	17.265588
mean	17.250132	mean	16.823745
std	293.847563	std	298.301160
std	116.586376	min	0.000000
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	0.000000	max	90750.000000
max	4528.899902	max	90750.000000
	Name: V135, dtype: float64	Name: V135, dtype: float64	
	Name: V135, dtype: float64		

The unique values in V135 feature is: 3725

V135 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V136 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	38.821196	mean	38.824830	
mean	38.720929			
std	451.808411	std	457.979479	
std	221.912050			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	90750.000000	max	90750.000000	
max	4528.899902			
Name: V136, dtype: float64		Name: V136, dtype: float64		
Name: V136, dtype: float64				

The unique values in V136 feature is: 4853

V136 feature has 314 null values, which holds 0.0005317167338368274 %

```
*****
*****
```

Analysis of V137 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590226.000000	count	569580.000000	
count	20646.000000			
mean	26.365090	mean	26.372703	
mean	26.155054			
std	348.332714	std	353.276445	
std	160.145334			

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000		min	0.000000	
min	0.000000				
25%	0.000000		25%	0.000000	
25%	0.000000				
50%	0.000000		50%	0.000000	
50%	0.000000				
75%	0.000000		75%	0.000000	
75%	0.000000				
max	90750.000000		max	90750.000000	
max	4528.899902				
Name: V137, dtype: float64			Name: V137, dtype: float64		
Name: V137, dtype: float64					
+-----+-----+-----+					

The unique values in V137 feature is: 4253

V137 feature has 314 null values, which holds 0.0005317167338368274 %

Analysis of V138 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature	
Discription of fraud transactions					
+-----+-----+-----+					
count	81945.000000		count	78295.000000	
count	3650.000000				
mean	0.036439		mean	0.032492	
mean	0.121096				
std	0.428490		std	0.414186	
std	0.659371				
min	0.000000		min	0.000000	
min	0.000000				
25%	0.000000		25%	0.000000	
25%	0.000000				
50%	0.000000		50%	0.000000	
50%	0.000000				
75%	0.000000		75%	0.000000	
75%	0.000000				
max	22.000000		max	22.000000	
max	8.000000				
Name: V138, dtype: float64			Name: V138, dtype: float64		
Name: V138, dtype: float64					
+-----+-----+-----+					

The unique values in V138 feature is: 24

V138 feature has 508595 null values, which holds 0.8612371727571375 %

```
*****
*****
```

Analysis of V139 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	1.073915	mean	1.031407	
mean	1.985753			
std	1.333924	std	1.260822	
std	2.231422			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	3.000000			
max	33.000000	max	33.000000	
max	24.000000			
Name: V139, dtype: float64		Name: V139, dtype: float64		
Name: V139, dtype: float64				

The unique values in V139 feature is: 35

V139 feature has 508595 null values, which holds 0.8612371727571375 %

```
*****
*****
```

Analysis of V140 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	1.125267	mean	1.062852	
mean	2.464110			
std	1.467850	std	1.327652	
std	2.947508			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	0.000000	25%	0.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	2.000000		
75%	1.000000	75%	1.000000
75%	3.000000		
max	33.000000	max	33.000000
max	27.000000		
Name: V140, dtype: float64		Name: V140, dtype: float64	
Name: V140, dtype: float64			

The unique values in V140 feature is: 35

V140 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V141 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	0.037696	mean	0.033719	
mean	0.123014			
std	0.215133	std	0.201249	
std	0.403386			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	5.000000	max	5.000000	
max	3.000000			
Name: V141, dtype: float64		Name: V141, dtype: float64		
Name: V141, dtype: float64				

The unique values in V141 feature is: 7

V141 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V142 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	0.048581	mean	0.043132	
mean	0.165479			
std	0.313888	std	0.287141	
std	0.655113			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	9.000000	max	6.000000	
max	9.000000			
Name: V142, dtype: float64		Name: V142, dtype: float64		
Name: V142, dtype: float64				

The unique values in V142 feature is: 11

V142 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V143 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81951.000000	count	78300.000000	
count	3651.000000			
mean	8.397006	mean	8.752248	
mean	0.778417			
std	55.267545	std	56.513468	
std	2.640451			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	0.000000		50%	0.000000	
50%		0.000000				
	75%	0.000000		75%	0.000000	
75%		1.000000				
	max	869.000000		max	869.000000	
max		47.000000				
	Name: V143, dtype: float64			Name: V143, dtype: float64		
Name: V143, dtype: float64						
<hr/>						
<hr/>						

The unique values in V143 feature is: 871

V143 feature has 508589 null values, which holds 0.8612270125647712 %

Analysis of V144 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81951.000000		count	78300.000000
count	3651.000000			
mean	3.708484		mean	3.826156
mean	1.184881			
std	10.485633		std	10.682965
std	3.701618			
min	0.000000		min	0.000000
min	0.000000			
25%	0.000000		25%	0.000000
25%	0.000000			
50%	0.000000		50%	0.000000
50%	0.000000			
75%	0.000000		75%	0.000000
75%	1.000000			
max	62.000000		max	62.000000
max	51.000000			
	Name: V144, dtype: float64		Name: V144, dtype: float64	
Name: V144, dtype: float64				
<hr/>				
<hr/>				

The unique values in V144 feature is: 64

V144 feature has 508589 null values, which holds 0.8612270125647712 %

Analysis of V145 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81951.000000	count	78300.000000	
count	3651.000000			
mean	22.112946	mean	22.974253	
mean	3.641194			
std	64.371860	std	65.556946	
std	22.013184			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	1.000000	75%	0.000000	
75%	1.500000			
max	297.000000	max	297.000000	
max	292.000000			
Name: V145, dtype: float64		Name: V145, dtype: float64		
Name: V145, dtype: float64				

The unique values in V145 feature is: 261

V145 feature has 508589 null values, which holds 0.8612270125647712 %

Analysis of V146 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	0.156276	mean	0.127020	
mean	0.783836			
std	0.682328	std	0.555728	
std	1.848351			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	0.000000	75%	0.000000
75%	1.000000		
max	24.000000	max	11.000000
max	24.000000		
Name: V146, dtype: float64		Name: V146, dtype: float64	
Name: V146, dtype: float64			

The unique values in V146 feature is: 26

V146 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V147 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	0.168601	mean	0.129242	
mean	1.012877			
std	0.750497	std	0.560253	
std	2.273225			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	26.000000	max	11.000000	
max	26.000000			
Name: V147, dtype: float64		Name: V147, dtype: float64		
Name: V147, dtype: float64				

The unique values in V147 feature is: 28

V147 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V148 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	0.765001	mean	0.736267	
mean	1.381370			
std	0.581425	std	0.456938	
std	1.647401			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			
max	20.000000	max	5.000000	
max	20.000000			
Name: V148, dtype: float64		Name: V148, dtype: float64		
Name: V148, dtype: float64				

The unique values in V148 feature is: 22

V148 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V149 feature***Five Number Summary***

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	0.775313	mean	0.738246	
mean	1.570411			
std	0.628191	std	0.459925	
std	1.913475			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	20.000000	max	5.000000
max	20.000000		
Name: V149, dtype: float64		Name: V149, dtype: float64	
Name: V149, dtype: float64			

The unique values in V149 feature is: 22
V149 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V150 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81951.000000	count	78300.000000	
count	3651.000000			
mean	277.598028	mean	289.056284	
mean	31.862230			
std	829.576922	std	844.935531	
std	271.047102			
min	1.000000	min	1.000000	
min	1.000000			
25%	1.000000	25%	1.000000	
50%	1.000000	50%	1.000000	
75%	1.000000	75%	1.000000	
75%	2.000000			
max	3389.000000	max	3389.000000	
max	3372.000000			
Name: V150, dtype: float64		Name: V150, dtype: float64		
Name: V150, dtype: float64				

The unique values in V150 feature is: 1997
V150 feature has 508589 null values, which holds 0.8612270125647712 %

Analysis of V151 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	81951.000000	count	78300.000000
count	3651.000000		
mean	6.460190	mean	6.660396
mean	2.166530		
std	15.232324	std	15.509543
std	5.474493		
min	1.000000	min	1.000000
min	1.000000		
25%	1.000000	25%	1.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	1.000000		
75%	1.000000	75%	1.000000
75%	1.000000		
max	57.000000	max	57.000000
max	55.000000		
	Name: V151, dtype: float64		Name: V151, dtype: float64
	Name: V151, dtype: float64		

The unique values in V151 feature is: 57

V151 feature has 508589 null values, which holds 0.8612270125647712 %

Analysis of V152 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	81951.000000	count	78300.000000
count	3651.000000		
mean	9.432710	mean	9.708033
mean	3.528075		
std	21.554486	std	21.895387
std	10.510815		
min	1.000000	min	1.000000
min	1.000000		
25%	1.000000	25%	1.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	1.000000		
75%	1.000000	75%	1.000000
75%	2.000000		
max	69.000000	max	69.000000
max	69.000000		

```
| Name: V152, dtype: float64 |           Name: V152, dtype: float64 |
| Name: V152, dtype: float64 |           |
+-----+-----+-----+
```

The unique values in V152 feature is: 40
V152 feature has 508589 null values, which holds 0.8612270125647712 %

```
*****
*****
```

Analysis of V153 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	0.753200	mean	0.732141	
mean	1.204932			
std	0.532649	std	0.449403	
std	1.350650			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	18.000000	max	4.000000	
max	18.000000			
Name: V153, dtype: float64		Name: V153, dtype: float64		
Name: V153, dtype: float64				

The unique values in V153 feature is: 20
V153 feature has 508595 null values, which holds 0.8612371727571375 %

```
*****
*****
```

Analysis of V154 feature

Five Number Summary

```
+-----+-----+-----+
```

1. self_case_study_1_IEEE_fraud_detection_EDA

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	81945.000000	count	78295.000000
mean	0.757423	mean	0.732984
std	0.546907	std	0.450330
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	18.000000	max	4.000000
	Name: V154, dtype: float64	Name: V154, dtype: float64	
	Name: V154, dtype: float64		

The unique values in V154 feature is: 20

V154 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V155 feature

Five Number Summary

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	81945.000000	count	78295.000000
mean	0.767381	mean	0.736905
std	0.602678	std	0.458811
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	24.000000	max	5.000000
	Name: V155, dtype: float64	Name: V155, dtype: float64	
	Name: V155, dtype: float64		

```
+-----+-----+
-----+-----+
```

The unique values in V155 feature is: 26
V155 feature has 508595 null values, which holds 0.8612371727571375 %

```
*****  
*****
```

Analysis of V156 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+-----+
| count      81945.000000 | count      78295.000000 |
count      3650.000000 | mean      0.738923 |
mean      1.604658 | std       0.461958 |
std       2.027660 | min       0.000000 |
min       0.000000 | max       5.000000 |
max       24.000000 | Name: V156, dtype: float64 |
Name: V156, dtype: float64 | Name: V156, dtype: float64 |
+-----+-----+-----+
```

The unique values in V156 feature is: 26
V156 feature has 508595 null values, which holds 0.8612371727571375 %

```
*****  
*****
```

Analysis of V157 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
```

count	81945.000000	count	78295.000000
count	3650.000000	mean	0.782987
mean	0.817573	std	0.546527
mean	1.559452	min	0.000000
std	0.682016	25%	0.000000
std	1.860259	50%	1.000000
min	0.000000	75%	1.000000
min	0.000000	max	7.000000
25%	0.000000	25%	0.000000
25%	1.000000	50%	1.000000
50%	1.000000	75%	1.000000
75%	1.000000	max	7.000000
75%	2.000000		
max	24.000000		
max	24.000000		
	Name: V157, dtype: float64	Name: V157, dtype: float64	
	Name: V157, dtype: float64		

The unique values in V157 feature is: 26

V157 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V158 feature

Five Number Summary

		Feature Description	Feature Description of legit transactions Feature Description of fraud transactions
count	81945.000000	count	78295.000000
count	3650.000000	mean	0.789374
mean	0.833461	std	0.557996
mean	1.779178	min	0.000000
std	0.734336	25%	0.000000
std	2.119609	50%	1.000000
min	0.000000	75%	1.000000
min	0.000000	max	7.000000
25%	0.000000		
25%	1.000000		
50%	1.000000		
50%	1.000000		
75%	1.000000		
75%	2.000000		
max	24.000000		
max	24.000000		
	Name: V158, dtype: float64	Name: V158, dtype: float64	
	Name: V158, dtype: float64		

The unique values in V158 feature is: 26
 V158 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V159 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81951.000000	count	78300.000000	
count	3651.000000			
mean	2719.299775	mean	2832.353551	
mean	294.728242			
std	8355.445049	std	8511.473919	
std	2687.570131			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	43551.929688			
Name: V159, dtype: float64		Name: V159, dtype: float64		
Name: V159, dtype: float64				

The unique values in V159 feature is: 6664
 V159 feature has 508589 null values, which holds 0.8612270125647712 %

Analysis of V160 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81951.000000	count	78300.000000	
count	3651.000000			
mean	2719.299775	mean	2832.353551	
mean	294.728242			
std	8355.445049	std	8511.473919	
std	2687.570131			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	43551.929688			
Name: V160, dtype: float64		Name: V160, dtype: float64		
Name: V160, dtype: float64				

1. self_case_study_1_IEEE_fraud_detection_EDA

count	81951.000000	count	78300.000000
count	3651.000000	mean	47453.181173
mean	6437.664702	mean	49365.667132
std	142076.069162	std	144620.034216
std	52765.489660	min	0.000000
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	0.000000	75%	0.000000
75%	0.000000	max	641511.437500
max	639717.437500	max	641511.437500
Name: V160, dtype: float64		Name: V160, dtype: float64	
Name: V160, dtype: float64			

The unique values in V160 feature is: 9622

V160 feature has 508589 null values, which holds 0.8612270125647712 %

Analysis of V161 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000	mean	4.337461	
mean	4.843938	mean	15.708219	
mean	15.708219	std	58.929757	
std	61.042699	std	58.780839	
min	0.000000	min	0.000000	
min	0.000000	25%	0.000000	
25%	0.000000	25%	0.000000	
25%	0.000000	50%	0.000000	
50%	0.000000	50%	0.000000	
50%	0.000000	75%	0.000000	
75%	0.000000	75%	0.000000	
75%	0.000000	max	3300.000000	
max	500.000000	max	3300.000000	
Name: V161, dtype: float64		Name: V161, dtype: float64		
Name: V161, dtype: float64				

The unique values in V161 feature is: 80
 V161 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V162 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			
mean	6.594661	mean	5.798320	
mean	23.676712			
std	69.195180	std	67.637239	
std	95.172398			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	3300.000000	max	3300.000000	
max	1170.000000			
Name: V162, dtype: float64		Name: V162, dtype: float64		
Name: V162, dtype: float64				

The unique values in V162 feature is: 186
 V162 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V163 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81945.000000	count	78295.000000	
count	3650.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	5.505955	mean	4.872220
mean	19.100000		
std	63.077887	std	62.374807
std	75.358863		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	0.000000		
max	3300.000000	max	3300.000000
max	770.000000		
Name: V163, dtype: float64		Name: V163, dtype: float64	
Name: V163, dtype: float64			

The unique values in V163 feature is: 107

V163 feature has 508595 null values, which holds 0.8612371727571375 %

Analysis of V164 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	81951.000000	count	78300.000000	
count	3651.000000			
mean	877.888928	mean	913.738450	
mean	109.053670			
std	6049.166505	std	6185.908288	
std	306.245568			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	100.000000			
max	93736.000000	max	93736.000000	
max	5450.000000			
Name: V164, dtype: float64		Name: V164, dtype: float64		
Name: V164, dtype: float64				

The unique values in V164 feature is: 1979

V164 feature has 508589 null values, which holds 0.8612270125647712 %

```
*****
*****
```

Analysis of V165 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	81951.000000	count	78300.000000
mean	2239.912219	mean	2331.180107
std	282.564747	std	8396.153254
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	150.000000	max	98476.000000
count	3651.000000	Name: V165, dtype: float64	Name: V165, dtype: float64
mean	17468.949219	Name: V165, dtype: float64	

The unique values in V165 feature is: 2548

V165 feature has 508589 null values, which holds 0.8612270125647712 %

```
*****
*****
```

Analysis of V166 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	81951.000000	count	78300.000000
mean	359.469437	mean	367.379528
std	189.828217	std	100.000000
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	150.000000	max	98476.000000
count	3651.000000	Name: V166, dtype: float64	Name: V166, dtype: float64
mean	17468.949219	Name: V166, dtype: float64	

1. self_case_study_1_IEEE_fraud_detection_EDA

std	1244.463270	std	1267.306208
std	536.808466	min	0.000000
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	175.000000	75%	0.000000
max	104060.000000	max	104060.000000
max	9000.000000		
	Name: V166, dtype: float64		Name: V166, dtype: float64
	Name: V166, dtype: float64		

The unique values in V166 feature is: 988

V166 feature has 508589 null values, which holds 0.8612270125647712 %

Analysis of V167 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	3.929514	mean	4.088443	
mean	2.078177			
std	42.200980	std	43.954364	
std	4.182894			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	1.000000			
75%	0.000000	75%	0.000000	
75%	2.000000			
max	872.000000	max	872.000000	
max	47.000000			
	Name: V167, dtype: float64		Name: V167, dtype: float64	
	Name: V167, dtype: float64			

The unique values in V167 feature is: 874

V167 feature has 450909 null values, which holds 0.763553696616656 %

```
*****
*****
```

Analysis of V168 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	139631.000000	128592.000000	
mean	5.859501	6.113133	
std	2.904973	56.262604	
min	54.032468	6.689211	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	1.000000	0.000000	
max	3.000000	964.000000	
	Name: V168, dtype: float64	Name: V168, dtype: float64	
	Name: V168, dtype: float64		

The unique values in V168 feature is: 966

V168 feature has 450909 null values, which holds 0.763553696616656 %

```
*****
*****
```

Analysis of V169 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	139819.000000	128749.000000	
mean	11070.000000	0.168053	
std	0.378229	0.149982	
min	0.904944	0.788091	
25%	1.752748		

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	19.000000	max	19.000000
max	18.000000		
Name: V169, dtype: float64		Name: V169, dtype: float64	
Name: V169, dtype: float64			

The unique values in V169 feature is: 21

V169 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V170 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			
mean	1.436779	mean	1.308515	
mean	2.928546			
std	1.751143	std	1.216551	
std	4.370712			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	2.000000			
75%	1.000000	75%	1.000000	
75%	3.000000			
max	48.000000	max	30.000000	
max	48.000000			
Name: V170, dtype: float64		Name: V170, dtype: float64		
Name: V170, dtype: float64				

The unique values in V170 feature is: 50

V170 feature has 450721 null values, which holds 0.7632353439225116 %

```
*****
*****
```

Analysis of V171 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			
mean	1.690908	mean	1.535701	
mean	3.496025			
std	2.444748	std	1.949572	
std	5.267626			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	2.000000			
75%	1.000000	75%	1.000000	
75%	4.000000			
max	61.000000	max	30.000000	
max	61.000000			
Name: V171, dtype: float64		Name: V171, dtype: float64		
Name: V171, dtype: float64				

The unique values in V171 feature is: 63

V171 feature has 450721 null values, which holds 0.7632353439225116 %

```
*****
*****
```

Analysis of V172 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	0.132012	mean	0.135599	
mean	0.090226			
std	0.924894	std	0.950226	
std	0.547949			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
75%	0.000000	max	31.000000
max	22.000000	max	31.000000
Name: V172, dtype: float64		Name: V172, dtype: float64	
Name: V172, dtype: float64			

The unique values in V172 feature is: 33

V172 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V173 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	0.055231	mean	0.056885	
mean	0.035963			
std	0.264439	std	0.269317	
std	0.197997			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	7.000000	max	7.000000	
max	3.000000			
Name: V173, dtype: float64		Name: V173, dtype: float64		
Name: V173, dtype: float64				

The unique values in V173 feature is: 9

V173 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V174 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			
mean	0.127794	mean	0.125989	
mean	0.14878			
std	0.379763	std	0.378054	
std	0.39852			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	8.000000	max	8.000000	
max	4.000000			
Name: V174, dtype: float64		Name: V174, dtype: float64		
Name: V174, dtype: float64				

The unique values in V174 feature is: 10

V174 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V175 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			
mean	0.212639	mean	0.213198	
mean	0.206143			
std	0.857133	std	0.868112	
std	0.717197			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	0.000000		50%	0.000000	
50%		0.000000				
	75%	0.000000		75%	0.000000	
75%		0.000000				
	max	14.000000		max	9.000000	
max		14.000000				
	Name: V175, dtype: float64			Name: V175, dtype: float64		
Name: V175, dtype: float64						
<hr/>						
<hr/>						

The unique values in V175 feature is: 16

V175 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V176 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	1.376979	mean	1.288890	
mean	2.403116			
std	1.829420	std	1.582096	
std	3.468886			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			
max	48.000000	max	48.000000	
max	48.000000			
Name: V176, dtype: float64		Name: V176, dtype: float64		
Name: V176, dtype: float64				
<hr/>				
<hr/>				

The unique values in V176 feature is: 50

V176 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V177 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	3.529245	mean	3.695253	
mean	1.595434			
std	41.310622	std	43.028239	
std	3.878491			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	2.000000			
max	861.000000	max	861.000000	
max	47.000000			
Name: V177, dtype: float64		Name: V177, dtype: float64		
Name: V177, dtype: float64				

The unique values in V177 feature is: 863

V177 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V178 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	6.649283	mean	6.985901	
mean	2.728055			
std	69.447364	std	72.325917	
std	7.235370			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	0.000000	75%	0.000000
75%	2.000000		
max	1235.000000	max	1235.000000
max	102.000000		
Name: V178, dtype: float64		Name: V178, dtype: float64	
Name: V178, dtype: float64			

The unique values in V178 feature is: 1237

V178 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V179 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	4.869012	mean	5.104283	
mean	2.128363			
std	50.489426	std	52.583847	
std	5.128527			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	2.000000			
max	920.000000	max	920.000000	
max	60.000000			
Name: V179, dtype: float64		Name: V179, dtype: float64		
Name: V179, dtype: float64				

The unique values in V179 feature is: 922

V179 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V180 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			
mean	0.924123	mean	0.940862	
mean	0.729449			
std	6.112630	std	6.271623	
std	3.797911			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	83.000000	max	83.000000	
max	68.000000			
Name: V180, dtype: float64		Name: V180, dtype: float64		
Name: V180, dtype: float64				

The unique values in V180 feature is: 85

V180 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V181 feature***Five Number Summary***

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	0.253332	mean	0.242962	
mean	0.374128			
std	1.248974	std	1.249050	
std	1.241782			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	24.000000	max	15.000000
max	24.000000		
Name: V181, dtype: float64		Name: V181, dtype: float64	
Name: V181, dtype: float64			

The unique values in V181 feature is: 26

V181 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V182 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	0.856243	mean	0.873398	
mean	0.656400			
std	5.792934	std	5.973453	
std	2.961885			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	0.000000	75%	0.000000	
75%	0.000000			
max	83.000000	max	83.000000	
max	49.000000			
Name: V182, dtype: float64		Name: V182, dtype: float64		
Name: V182, dtype: float64				

The unique values in V182 feature is: 85

V182 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V183 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	139631.000000	count	128592.000000
count	11039.000000	mean	0.485256
mean	0.484692	mean	0.478123
std	2.946982	std	3.034699
std	1.603992	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	0.000000	max	41.000000
max	31.000000	max	41.000000
	Name: V183, dtype: float64	Name: V183, dtype: float64	
	Name: V183, dtype: float64		

The unique values in V183 feature is: 43

V183 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V184 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions	
count	139819.000000	count	128749.000000
count	11070.000000	mean	0.113011
mean	0.132364	mean	0.357453
std	0.543699	std	0.478478
std	1.007989	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	0.000000	max	11.000000
max	16.000000	max	16.000000

```
| Name: V184, dtype: float64 |           Name: V184, dtype: float64 |
| Name: V184, dtype: float64 |           |
+-----+-----+-----+
```

The unique values in V184 feature is: 18

V184 feature has 450721 null values, which holds 0.7632353439225116 %

```
*****
*****
```

Analysis of V185 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	139819.000000	count	128749.000000	count	11070.000000
mean	0.173903	mean	0.151489	mean	0.434598
std	0.699733	std	0.582022	std	1.473350
min	0.000000	min	0.000000	min	0.000000
25%	0.000000	25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000	75%	0.000000
max	31.000000	max	11.000000	max	31.000000
	Name: V185, dtype: float64		Name: V185, dtype: float64		Name: V185, dtype: float64
	Name: V185, dtype: float64				

The unique values in V185 feature is: 33

V185 feature has 450721 null values, which holds 0.7632353439225116 %

```
*****
*****
```

Analysis of V186 feature

Five Number Summary

```
+-----+-----+-----+
```

1. self_case_study_1_IEEE_fraud_detection_EDA

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	139631.000000	count	128592.000000
mean	1.148799	mean	1.095418
std	1.223529	std	1.134161
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	38.000000	max	38.000000
	Name: V186, dtype: float64	Name: V186, dtype: float64	
	Name: V186, dtype: float64		

The unique values in V186 feature is: 40

V186 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V187 feature

Five Number Summary

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	139631.000000	count	128592.000000
mean	1.843752	mean	1.773586
std	9.884390	std	9.910548
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	218.000000	max	217.000000
	Name: V187, dtype: float64	Name: V187, dtype: float64	
	Name: V187, dtype: float64		

```
+-----+-----+
-----+-----+
```

The unique values in V187 feature is: 216
V187 feature has 450909 null values, which holds 0.763553696616656 %

```
*****  
*****
```

Analysis of V188 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description      | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+-----+
| count      139819.000000 |           count      128749.000000 |
count      11070.000000 |           mean       0.955005 |
mean       1.709666 |           std        0.328935 |
std        0.671207 |           min       0.000000 |
min        0.000000 |           25%       1.000000 |
25%        1.000000 |           50%       1.000000 |
50%        1.000000 |           75%       1.000000 |
75%        2.000000 |           max       7.000000 |
max        30.000000 |           Name: V188, dtype: float64 |
Name: V188, dtype: float64 |           Name: V188, dtype: float64 |
+-----+-----+-----+
```

The unique values in V188 feature is: 32
V188 feature has 450721 null values, which holds 0.7632353439225116 %

```
*****  
*****
```

Analysis of V189 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description      | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+-----+
```

count	139819.000000		count	128749.000000	
count	11070.000000				
mean	1.038314		mean	0.966687	
mean	1.871364				
std	0.792528		std	0.432240	
std	2.237640				
min	0.000000		min	0.000000	
min	0.000000				
25%	1.000000		25%	1.000000	
25%	1.000000				
50%	1.000000		50%	1.000000	
50%	1.000000				
75%	1.000000		75%	1.000000	
75%	2.000000				
max	30.000000		max	13.000000	
max	30.000000				
Name: V189, dtype: float64			Name: V189, dtype: float64		
Name: V189, dtype: float64					

The unique values in V189 feature is: 32

v189 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V190 feature

Five Number Summary

Feature Description Description of fraud transactions		Feature Description of legit transactions		Feature
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	1.213083	mean	1.124106	
mean	2.249570			
std	1.480318	std	1.212505	
std	3.070257			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			
max	42.000000	max	40.000000	
max	42.000000			
Name: V190, dtype: float64		Name: V190, dtype: float64		
Name: V190, dtype: float64				

The unique values in V190 feature is: 44
 V190 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V191 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	139631.000000	count	128592.000000
mean	1.058984	mean	1.048028
std	0.677527	std	0.668029
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	21.000000	max	21.000000
	Name: V191, dtype: float64	Name: V191, dtype: float64	
	Name: V191, dtype: float64		

The unique values in V191 feature is: 23
 V191 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V192 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	139631.000000	count	128592.000000
mean	1.058984	mean	1.048028
std	0.677527	std	0.668029
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	21.000000	max	21.000000
	Name: V192, dtype: float64	Name: V192, dtype: float64	
	Name: V192, dtype: float64		

1. self_case_study_1_IEEE_fraud_detection_EDA

count	139631.000000	count	128592.000000
count	11039.000000	mean	1.223194
mean	1.237361	mean	1.402392
mean	2.635961	std	2.511878
std	2.645872	std	0.000000
std	0.000000	min	1.000000
min	1.000000	min	1.000000
min	1.000000	25%	1.000000
25%	1.000000	25%	1.000000
25%	1.000000	50%	1.000000
50%	1.000000	50%	1.000000
50%	1.000000	75%	1.000000
75%	1.000000	75%	1.000000
75%	1.000000	max	44.000000
max	43.000000	max	44.000000
max	44.000000	Name: V192, dtype: float64	Name: V192, dtype: float64
Name: V192, dtype: float64	Name: V192, dtype: float64	+-----+	+-----+
+-----+	+-----+	+-----+	+-----+

The unique values in V192 feature is: 46

V192 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V193 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	count
count	11039.000000	mean	1.138757	128592.000000
mean	1.149845	mean	1.279011	mean
mean	1.739193	std	1.393129	1.714792
std	0.000000	std	0.000000	1.714792
min	0.000000	min	0.000000	1.393129
25%	1.000000	25%	1.000000	0.000000
25%	1.000000	50%	1.000000	1.000000
50%	1.000000	50%	1.000000	1.000000
50%	1.000000	75%	1.000000	1.000000
75%	1.000000	75%	1.000000	1.000000
75%	1.000000	max	37.000000	1.000000
max	30.000000	max	37.000000	30.000000
max	30.000000	Name: V193, dtype: float64	Name: V193, dtype: float64	Name: V193, dtype: float64
Name: V193, dtype: float64	Name: V193, dtype: float64	+-----+	+-----+	+-----+
+-----+	+-----+	+-----+	+-----+	+-----+

The unique values in V193 feature is: 39
 V193 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V194 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			
mean	0.945773	mean	0.932644	
mean	1.098464			
std	0.304158	std	0.264245	
std	0.575392			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	7.000000	max	3.000000	
max	7.000000			
Name: V194, dtype: float64		Name: V194, dtype: float64		
Name: V194, dtype: float64				

The unique values in V194 feature is: 9
 V194 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V195 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	0.954227	mean	0.937996
mean	1.142999	std	0.306818
std	0.376420	std	0.306818
std	0.810015	min	0.000000
min	0.000000	25%	1.000000
25%	1.000000	50%	1.000000
50%	1.000000	75%	1.000000
75%	1.000000	max	8.000000
max	16.000000	max	8.000000
	Name: V195, dtype: float64		Name: V195, dtype: float64
	Name: V195, dtype: float64		

The unique values in V195 feature is: 18

V195 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V196 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	1.084580	mean	1.072182	
mean	1.229006	std	1.105736	
std	1.105834	std	1.105736	
std	1.096746	min	0.000000	
min	0.000000	min	0.000000	
25%	1.000000	25%	1.000000	
25%	1.000000	50%	1.000000	
50%	1.000000	50%	1.000000	
75%	1.000000	75%	1.000000	
75%	1.000000	max	38.000000	
max	38.000000	max	38.000000	
	Name: V196, dtype: float64		Name: V196, dtype: float64	
	Name: V196, dtype: float64			

The unique values in V196 feature is: 40

V196 feature has 450909 null values, which holds 0.763553696616656 %

```
*****
*****
```

Analysis of V197 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			
mean	0.949385	mean	0.934174	
mean	1.126287			
std	0.336871	std	0.268115	
std	0.750550			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	14.000000	max	4.000000	
max	14.000000			
Name: V197, dtype: float64		Name: V197, dtype: float64		
Name: V197, dtype: float64				

The unique values in V197 feature is: 16

V197 feature has 450721 null values, which holds 0.7632353439225116 %

```
*****
*****
```

Analysis of V198 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139819.000000	count	128749.000000	
count	11070.000000			
mean	0.961951	mean	0.941724	
mean	1.197200			

1. self_case_study_1_IEEE_fraud_detection_EDA

std	0.493065	std	0.384678
std	1.135603	min	0.000000
min	0.000000	25%	1.000000
25%	1.000000	50%	1.000000
50%	1.000000	75%	1.000000
75%	1.000000	max	14.000000
max	21.000000	Name: V198, dtype: float64	Name: V198, dtype: float64
	21.000000	Name: V198, dtype: float64	
-----+-----+-----+-----			

The unique values in V198 feature is: 23

V198 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V199 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	139631.000000	count	128592.000000
count	11039.000000	mean	1.157584
mean	1.270749	std	1.297311
std	2.589003	min	0.000000
min	1.671863	25%	0.000000
25%	3.723457	50%	1.000000
50%	1.000000	75%	1.000000
75%	1.000000	max	43.000000
max	3.000000	Name: V199, dtype: float64	Name: V199, dtype: float64
	45.000000	Name: V199, dtype: float64	
-----+-----+-----			

The unique values in V199 feature is: 47

V199 feature has 450909 null values, which holds 0.763553696616656 %

```
*****
*****
```

Analysis of V200 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	139819.000000	128749.000000	
count	11070.000000		
mean	1.119977	1.002773	
mean	2.483107		
std	1.253853	0.442942	
std	3.944418		
min	0.000000	0.000000	
min	0.000000		
25%	1.000000	1.000000	
25%	1.000000		
50%	1.000000	1.000000	
50%	1.000000		
75%	1.000000	1.000000	
75%	3.000000		
max	45.000000	12.000000	
max	45.000000		
	Name: V200, dtype: float64	Name: V200, dtype: float64	
	Name: V200, dtype: float64		

The unique values in V200 feature is: 47

V200 feature has 450721 null values, which holds 0.7632353439225116 %

```
*****
*****
```

Analysis of V201 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	139819.000000	128749.000000	
count	11070.000000		
mean	1.159106	1.022664	
mean	2.745980		
std	1.418618	0.567060	
std	4.352642		

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000		min	0.000000	
min	0.000000				
25%	1.000000		25%	1.000000	
25%	1.000000				
50%	1.000000		50%	1.000000	
50%	1.000000				
75%	1.000000		75%	1.000000	
75%	3.000000				
max	55.000000		max	15.000000	
max	55.000000				
Name: V201, dtype: float64		Name: V201, dtype: float64			
Name: V201, dtype: float64					
+-----+-----+-----+					

The unique values in V201 feature is: 57

V201 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V202 feature

Five Number Summary

	Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	139631.000000		count	128592.000000		
count	11039.000000					
mean	444.147142		mean	468.886764		
mean	155.958232					
std	4683.828419		std	4878.506934		
std	403.621902					
min	0.000000		min	0.000000		
min	0.000000					
25%	0.000000		25%	0.000000		
25%	0.000000					
50%	0.000000		50%	0.000000		
50%	21.388700					
75%	0.000000		75%	0.000000		
75%	150.000000					
max	104060.000000		max	104060.000000		
max	10900.000000					
Name: V202, dtype: float64		Name: V202, dtype: float64				
Name: V202, dtype: float64						
+-----+-----+-----+						

The unique values in V202 feature is: 10971

V202 feature has 450909 null values, which holds 0.763553696616656 %

```
*****
*****
```

Analysis of V203 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	1078.327538	mean	1131.530069	
mean	458.577578			
std	9105.607991	std	9428.292395	
std	3581.819720			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	47.585400			
75%	30.924400	75%	19.757299	
75%	237.288399			
max	139777.000000	max	139777.000000	
max	81450.000000			
Name: V203, dtype: float64		Name: V203, dtype: float64		
Name: V203, dtype: float64				

The unique values in V203 feature is: 14952

V203 feature has 450909 null values, which holds 0.763553696616656 %

```
*****
*****
```

Analysis of V204 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	686.956931	mean	723.728644	
mean	258.607617			
std	6048.980716	std	6291.648851	
std	1227.476243			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	38.011600	50%	0.000000
75%	20.000000	75%	0.000000
75%	200.000000	max	104060.000000
max	104060.000000	max	104060.000000
max	37850.000000	Name: V204, dtype: float64	Name: V204, dtype: float64
		Name: V204, dtype: float64	

The unique values in V204 feature is: 12859

V204 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V205 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	18.060417	mean	18.833240	
mean	9.057890			
std	266.545451	std	275.648287	
std	116.047755			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	5050.000000			
Name: V205, dtype: float64		Name: V205, dtype: float64		
Name: V205, dtype: float64				

The unique values in V205 feature is: 2241

V205 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V206 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	6.189360	mean	6.508866	
mean	2.467472			
std	191.474014	std	199.382949	
std	25.255635			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	1350.000000			
Name: V206, dtype: float64		Name: V206, dtype: float64		
Name: V206, dtype: float64				

The unique values in V206 feature is: 1781

V206 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V207 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	72.284098	mean	75.651962	
mean	33.052247			
std	925.676660	std	954.372372	
std	476.216252			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	0.000000		50%	0.000000	
50%		0.000000				
	75%	0.000000		75%	0.000000	
75%		0.000000				
	max	55125.000000		max	55125.000000	
max		20075.000000				
	Name: V207, dtype: float64			Name: V207, dtype: float64		
Name: V207, dtype: float64						
<hr/>						
<hr/>						

The unique values in V207 feature is: 3247

V207 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V208 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	139819.000000		count	128749.000000	
count	11070.000000				
mean	8.888193		mean	8.729766	
mean	10.730773				
std	59.047125		std	60.022465	
std	46.177884				
min	0.000000		min	0.000000	
min	0.000000				
25%	0.000000		25%	0.000000	
25%	0.000000				
50%	0.000000		50%	0.000000	
50%	0.000000				
75%	0.000000		75%	0.000000	
75%	0.000000				
max	3300.000000		max	3300.000000	
max	835.000000				
Name: V208, dtype: float64			Name: V208, dtype: float64		
Name: V208, dtype: float64					
<hr/>					
<hr/>					

The unique values in V208 feature is: 2553

V208 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V209 feature

Five Number Summary

	Feature Description of legit transactions	Feature Description of fraud transactions	
count	139819.000000	count	128749.000000
count	11070.000000		
mean	35.001872	mean	36.386541
mean	18.897558		
std	250.800822	std	258.627612
std	127.469431		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	0.000000		
max	8050.000000	max	8050.000000
max	4725.000000		
Name: V209, dtype: float64		Name: V209, dtype: float64	
Name: V209, dtype: float64			

The unique values in V209 feature is: 3452

V209 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V210 feature

Five Number Summary

	Feature Description of legit transactions	Feature Description of fraud transactions	
count	139819.000000	count	128749.000000
count	11070.000000		
mean	14.391835	mean	14.408390
mean	14.199290		
std	86.523724	std	87.555739
std	73.466863		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	0.000000	75%	0.000000
75%	0.000000		
max	3300.000000	max	3300.000000
max	1825.000000		
Name: V210, dtype: float64		Name: V210, dtype: float64	
Name: V210, dtype: float64			

The unique values in V210 feature is: 2837
V210 feature has 450721 null values, which holds 0.7632353439225116 %

Analysis of V211 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	139631.000000	count	128592.000000
count	11039.000000		
mean	385.137037	mean	409.851898
mean	97.236561		
std	4541.837915	std	4731.337313
std	261.397582		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	86.883999		
max	92888.000000	max	92888.000000
max	5450.000000		
Name: V211, dtype: float64		Name: V211, dtype: float64	
Name: V211, dtype: float64			

The unique values in V211 feature is: 7625
V211 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V212 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	765.988339	mean	812.414172	
mean	225.179319			
std	7496.120737	std	7798.955699	
std	1385.125382			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	0.000000	75%	0.000000	
75%	141.404007			
max	129006.000000	max	129006.000000	
max	34625.000000			
Name: V212, dtype: float64		Name: V212, dtype: float64		
Name: V212, dtype: float64				

The unique values in V212 feature is: 8869

V212 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V213 feature***Five Number Summary***

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	536.302802	mean	569.329910	
mean	151.573946			
std	5471.664736	std	5697.325996	
std	647.084991			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	119.103798			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	97628.000000	max	97628.000000
max	16950.000000		
Name: V213, dtype: float64		Name: V213, dtype: float64	
Name: V213, dtype: float64			

The unique values in V213 feature is: 8318

V213 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V214 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	38.437547	mean	37.624827	
mean	47.904827			
std	571.834283	std	593.016828	
std	198.676843			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	0.000000	75%	0.000000	
75%	0.000000			
max	104060.000000	max	104060.000000	
max	4500.000000			
Name: V214, dtype: float64		Name: V214, dtype: float64		
Name: V214, dtype: float64				

The unique values in V214 feature is: 2283

V214 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V215 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions		
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	133.208217	mean	134.671741	
mean	116.159801			
std	1040.453748	std	1059.481445	
std	785.388033			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	104060.000000	max	104060.000000	
max	13725.000000			
Name: V215, dtype: float64		Name: V215, dtype: float64		
Name: V215, dtype: float64				

The unique values in V215 feature is: 2748

V215 feature has 450909 null values, which holds 0.763553696616656 %

Analysis of V216 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions		
count	139631.000000	count	128592.000000	
count	11039.000000			
mean	71.107143	mean	71.229102	
mean	69.686454			
std	680.267625	std	703.167916	
std	306.143918			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	104060.000000	max	104060.000000	
max	6100.000000			

```
| Name: V216, dtype: float64 |           Name: V216, dtype: float64 |
| Name: V216, dtype: float64 |           |
+-----+-----+-----+
```

The unique values in V216 feature is: 2533

V216 feature has 450909 null values, which holds 0.763553696616656 %

```
*****
*****
```

Analysis of V217 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.054619	mean	0.927352	
mean	2.556024			
std	9.547354	std	9.846424	
std	4.503774			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	1.000000			
75%	0.000000	75%	0.000000	
75%	3.000000			
max	303.000000	max	303.000000	
max	53.000000			
Name: V217, dtype: float64		Name: V217, dtype: float64		
Name: V217, dtype: float64				

The unique values in V217 feature is: 305

V217 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V218 feature

Five Number Summary

1. self_case_study_1_IEEE_fraud_detection_EDA

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	130430.000000	count	120238.000000
mean	1.725784	mean	1.533018
std	13.919876	std	14.342145
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	2.000000	75%	1.000000
max	400.000000	max	400.000000
	max		
	Name: V218, dtype: float64	Name: V218, dtype: float64	
	Name: V218, dtype: float64		

The unique values in V218 feature is: 402

V218 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V219 feature

Five Number Summary

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	130430.000000	count	120238.000000
mean	1.367132	mean	1.205684
std	12.044032	std	12.432478
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	1.000000	75%	0.000000
max	378.000000	max	378.000000
	max		
	Name: V219, dtype: float64	Name: V219, dtype: float64	
	Name: V219, dtype: float64		

```
+-----+-----+
-----+-----+
```

The unique values in V219 feature is: 380
V219 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****  
*****
```

Analysis of V220 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+-----+
| count      141416.000000 | | count      130434.000000 |
count      10982.000000 | | mean       0.147186 |
mean       0.434165 | | std        0.860589 |
std        2.017091 | | min        0.000000 |
min        0.000000 | | max        24.000000 |
max        25.000000 | | Name: V220, dtype: float64 |
Name: V220, dtype: float64 | | Name: V220, dtype: float64 |
+-----+-----+-----+
```

The unique values in V220 feature is: 27
V220 feature has 449124 null values, which holds 0.7605310393876791 %

```
*****  
*****
```

Analysis of V221 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
```

count	141416.000000	count	130434.000000
count	10982.000000	mean	1.169963
mean	1.274912	std	2.514822
mean	2.521399	min	0.000000
std	2.728792	25%	1.000000
std	4.369119	50%	1.000000
min	0.000000	75%	1.000000
min	0.000000	max	384.000000
25%	1.000000	25%	1.000000
25%	0.000000	50%	1.000000
50%	1.000000	75%	1.000000
75%	3.000000	max	384.000000
max	384.000000	Name: V221, dtype: float64	Name: V221, dtype: float64
max	84.000000	Name: V221, dtype: float64	

The unique values in V221 feature is: 78

V221 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V222 feature

Five Number Summary

		Feature Description of legit transactions	Feature Description of fraud transactions
count	141416.000000	count	130434.000000
count	10982.000000	mean	1.218333
mean	1.359005	std	2.603629
mean	3.029776	min	0.000000
std	2.913772	25%	1.000000
std	5.078343	50%	1.000000
min	0.000000	75%	1.000000
min	0.000000	max	384.000000
25%	1.000000	Name: V222, dtype: float64	Name: V222, dtype: float64
25%	0.000000		
50%	1.000000		
50%	1.000000		
75%	1.000000		
75%	4.000000		
max	384.000000		
max	86.000000		

The unique values in V222 feature is: 77
 V222 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V223 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	130430.000000	120238.000000	
mean	0.093153	0.093024	
std	0.094682	0.406483	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	16.000000	16.000000	
Name:	V223, dtype: float64	Name: V223, dtype: float64	
	Name: V223, dtype: float64		

The unique values in V223 feature is: 18
 V223 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V224 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	130430.000000	120238.000000	
mean	0.093153	0.093024	
std	0.094682	0.406483	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	16.000000	16.000000	
Name:	V224, dtype: float64	Name: V224, dtype: float64	
	Name: V224, dtype: float64		

1. self_case_study_1_IEEE_fraud_detection_EDA

count 130430.000000	count 120238.000000
count 10192.000000	
mean 0.394748	mean 0.393594
mean 0.408359	
std 3.443310	std 3.509631
std 2.533156	
min 0.000000	min 0.000000
min 0.000000	
25% 0.000000	25% 0.000000
25% 0.000000	
50% 0.000000	50% 0.000000
50% 0.000000	
75% 0.000000	75% 0.000000
75% 0.000000	
max 144.000000	max 144.000000
max 142.000000	
Name: V224, dtype: float64	Name: V224, dtype: float64
Name: V224, dtype: float64	

The unique values in V224 feature is: 80

V224 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V225 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count 130430.000000	count 120238.000000	count 120238.000000	count 120238.000000	
count 10192.000000		mean 0.188776	mean 0.188776	
mean 0.191758	mean 0.188776	mean 0.188776	mean 0.188776	
mean 0.226943		std 1.318557	std 1.318557	
std 1.290820	std 1.318557	std 1.318557	std 1.318557	
std 0.900620		min 0.000000	min 0.000000	
min 0.000000	min 0.000000	min 0.000000	min 0.000000	
min 0.000000		25% 0.000000	25% 0.000000	
25% 0.000000	25% 0.000000	25% 0.000000	25% 0.000000	
25% 0.000000				
50% 0.000000	50% 0.000000	50% 0.000000	50% 0.000000	
50% 0.000000				
75% 0.000000	75% 0.000000	75% 0.000000	75% 0.000000	
75% 0.000000				
max 51.000000	max 51.000000	max 51.000000	max 51.000000	
max 32.000000				
Name: V225, dtype: float64	Name: V225, dtype: float64	Name: V225, dtype: float64	Name: V225, dtype: float64	
Name: V225, dtype: float64				

The unique values in V225 feature is: 36
 V225 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V226 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	0.247612	mean	0.245788	
mean	0.269133			
std	2.756449	std	2.849110	
std	1.212436			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	242.000000	max	242.000000	
max	23.000000			
Name: V226, dtype: float64		Name: V226, dtype: float64		
Name: V226, dtype: float64				

The unique values in V226 feature is: 82
 V226 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V227 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	141416.000000	count	130434.000000	
count	10982.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	0.147975	mean	0.137870
mean	0.267984	std	2.070691
std	2.049363	std	1.772261
min	0.000000	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
75%	0.000000	max	360.000000
max	360.000000	max	79.000000
Name: V227, dtype: float64		Name: V227, dtype: float64	
Name: V227, dtype: float64			

The unique values in V227 feature is: 51

V227 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V228 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.351928	mean	1.241579	
mean	2.653748	std	0.951216	
std	1.409722	std	3.594541	
std	3.594541	min	0.000000	
min	0.000000	min	0.000000	
25%	1.000000	25%	1.000000	
25%	1.000000	50%	1.000000	
50%	1.000000	50%	1.000000	
75%	1.000000	75%	1.000000	
75%	3.000000	max	29.000000	
max	54.000000	max	54.000000	
Name: V228, dtype: float64		Name: V228, dtype: float64		
Name: V228, dtype: float64				

The unique values in V228 feature is: 56

V228 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V229 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	130430.000000	count	120238.000000
mean	1.638710	mean	1.466275
std	3.672979	std	4.163011
min	4.316026	min	5.426587
25%	0.000000	25%	0.000000
50%	1.000000	50%	1.000000
75%	2.000000	75%	1.000000
max	4.000000	max	1.000000
count	176.000000	count	128.000000
Name: V229, dtype: float64		Name: V229, dtype: float64	
Name: V229, dtype: float64			

The unique values in V229 feature is: 92

V229 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V230 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	130430.000000	count	120238.000000
mean	10192.000000	mean	1.465476
std	3.155122	std	1.322253
min	4.165476	min	5.155122
25%	0.000000	25%	0.000000
50%	1.000000	50%	1.000000
75%	2.000000	75%	1.000000
max	4.000000	max	1.000000
Name: V230, dtype: float64		Name: V230, dtype: float64	
Name: V230, dtype: float64			

1. self_case_study_1_IEEE_fraud_detection_EDA

std	2.122783	std	1.757632
std	4.257576	min	0.000000
min	0.000000	25%	1.000000
25%	1.000000	50%	1.000000
50%	2.000000	75%	1.000000
75%	3.000000	max	65.000000
max	54.000000	Name: V230, dtype: float64	Name: V230, dtype: float64
		Name: V230, dtype: float64	

The unique values in V230 feature is: 67

V230 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V231 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	0.765506	mean	0.658793	
mean	2.024431			
std	8.877119	std	9.153875	
std	4.267971			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	2.000000			
max	293.000000	max	293.000000	
max	51.000000			
Name: V231, dtype: float64		Name: V231, dtype: float64		
Name: V231, dtype: float64				

The unique values in V231 feature is: 295

V231 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V232 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	130430.000000	120238.000000	
count	10192.000000		
mean	1.005597	0.839310	
mean	2.967327		
std	10.460253	10.738334	
std	5.975140		
min	0.000000	0.000000	
min	0.000000		
25%	0.000000	0.000000	
25%	0.000000		
50%	0.000000	0.000000	
50%	1.000000		
75%	0.000000	0.000000	
75%	3.000000		
max	337.000000	337.000000	
max	56.000000		
Name: V232, dtype: float64		Name: V232, dtype: float64	
Name: V232, dtype: float64			

The unique values in V232 feature is: 339

V232 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V233 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	130430.000000	120238.000000	
count	10192.000000		
mean	0.908388	0.771786	
mean	2.519918		
std	10.149884	10.457271	
std	5.047551		

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
50%	1.000000	75%	0.000000
75%	3.000000	75%	0.000000
max	332.000000	max	332.000000
max	51.000000		
Name: V233, dtype: float64		Name: V233, dtype: float64	
Name: V233, dtype: float64		+-----+-----+	
+-----+-----+		+-----+-----+	

The unique values in V233 feature is: 334

V233 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V234 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	141416.000000	count	130434.000000	
count	10982.000000			
mean	2.092847	mean	2.156447	
mean	1.337461			
std	11.323036	std	11.572216	
std	7.735587			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	121.000000	max	121.000000	
max	120.000000			
Name: V234, dtype: float64		Name: V234, dtype: float64		
Name: V234, dtype: float64		+-----+-----+		
+-----+-----+		+-----+-----+		

The unique values in V234 feature is: 123

V234 feature has 449124 null values, which holds 0.7605310393876791 %

```
*****
*****
```

Analysis of V235 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	0.184106	mean	0.164083	
mean	0.420330			
std	0.902639	std	0.865420	
std	1.237212			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	23.000000	max	15.000000	
max	23.000000			
Name: V235, dtype: float64		Name: V235, dtype: float64		
Name: V235, dtype: float64				

The unique values in V235 feature is: 25

V235 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V236 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	0.307667	mean	0.282614	
mean	0.603218			
std	2.152332	std	2.196789	
std	1.502342			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
75%	1.000000		
max	45.000000	max	45.000000
max	23.000000		
Name: V236, dtype: float64		Name: V236, dtype: float64	
Name: V236, dtype: float64			

The unique values in V236 feature is: 47

V236 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V237 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	0.253500	mean	0.232148	
mean	0.505396			
std	1.716634	std	1.741805	
std	1.360593			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	39.000000	max	39.000000	
max	23.000000			
Name: V237, dtype: float64		Name: V237, dtype: float64		
Name: V237, dtype: float64				

The unique values in V237 feature is: 41

V237 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V238 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	141416.000000	count	130434.000000	
count	10982.000000			
mean	0.127956	mean	0.108361	
mean	0.360681			
std	0.574657	std	0.503668	
std	1.086645			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	23.000000	max	23.000000	
max	23.000000			
Name: V238, dtype: float64		Name: V238, dtype: float64		
Name: V238, dtype: float64				

The unique values in V238 feature is: 25

V238 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V239 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	141416.000000	count	130434.000000	
count	10982.000000			
mean	0.136965	mean	0.111681	
mean	0.437261			
std	0.616260	std	0.512448	
std	1.293764			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
75%	0.000000	max	23.000000
max	23.000000	max	23.000000
	Name: V239, dtype: float64		Name: V239, dtype: float64
	Name: V239, dtype: float64		

The unique values in V239 feature is: 25

V239 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V240 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.000997	mean	1.000956	
mean	1.001472			
std	0.049522	std	0.050191	
std	0.040816			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	7.000000	max	7.000000	
max	2.000000			
	Name: V240, dtype: float64		Name: V240, dtype: float64	
	Name: V240, dtype: float64			

The unique values in V240 feature is: 7

V240 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V241 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature
count	130430.000000	count	120238.000000
count	10192.000000		
mean	1.000238	mean	1.000225
mean	1.000392		
std	0.022663	std	0.022523
std	0.024261		
min	0.000000	min	0.000000
min	0.000000		
25%	1.000000	25%	1.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	1.000000		
75%	1.000000	75%	1.000000
75%	1.000000		
max	5.000000	max	5.000000
max	2.000000		
	Name: V241, dtype: float64	Name: V241, dtype: float64	
	Name: V241, dtype: float64		

The unique values in V241 feature is: 6

V241 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V242 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature
count	130430.000000	count	120238.000000
count	10192.000000		
mean	1.113463	mean	1.044162
mean	1.931024		
std	0.660110	std	0.345990
std	1.854551		
min	0.000000	min	0.000000
min	0.000000		
25%	1.000000	25%	1.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	1.000000		

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	1.000000	75%	1.000000
75%	2.000000		
max	20.000000	max	18.000000
max	20.000000		
Name: V242, dtype: float64		Name: V242, dtype: float64	
Name: V242, dtype: float64			

The unique values in V242 feature is: 22

V242 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V243 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	130430.000000	count	120238.000000
count	10192.000000		
mean	1.178387	mean	1.092758
mean	2.188579		
std	1.398953	std	1.258508
std	2.292076		
min	0.000000	min	0.000000
min	0.000000		
25%	1.000000	25%	1.000000
25%	1.000000		
50%	1.000000	50%	1.000000
50%	1.000000		
75%	1.000000	75%	1.000000
75%	2.000000		
max	57.000000	max	57.000000
max	34.000000		
Name: V243, dtype: float64		Name: V243, dtype: float64	
Name: V243, dtype: float64			

The unique values in V243 feature is: 44

V243 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V244 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.118562	mean	1.044478	
mean	1.992543			
std	0.698807	std	0.353183	
std	1.987337			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	2.000000	75%	1.000000	
75%	2.000000			
max	22.000000	max	18.000000	
max	22.000000			
Name: V244, dtype: float64		Name: V244, dtype: float64		
Name: V244, dtype: float64				

The unique values in V244 feature is: 24

V244 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V245 feature***Five Number Summary***

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	141416.000000	count	130434.000000	
count	10982.000000			
mean	0.876619	mean	0.828365	
mean	1.449736			
std	1.821065	std	1.758696	
std	2.369050			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	0.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	262.000000	max	262.000000
max	67.000000		
Name: V245, dtype: float64		Name: V245, dtype: float64	
Name: V245, dtype: float64			

The unique values in V245 feature is: 59
V245 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V246 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.183723	mean	1.072598	
mean	2.494702			
std	1.040363	std	0.425933	
std	3.138070			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	3.000000			
max	45.000000	max	18.000000	
max	45.000000			
Name: V246, dtype: float64		Name: V246, dtype: float64		
Name: V246, dtype: float64				

The unique values in V246 feature is: 47
V246 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V247 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.025355	mean	1.010621	
mean	1.199176			
std	0.281604	std	0.191675	
std	0.740734			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	18.000000	max	12.000000	
max	18.000000			
Name: V247, dtype: float64		Name: V247, dtype: float64		
Name: V247, dtype: float64				

The unique values in V247 feature is: 20

V247 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V248 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.069493	mean	1.049186	
mean	1.309066			
std	0.888341	std	0.882866	
std	0.917264			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	36.000000	max	36.000000	
max	18.000000			

```
| Name: V248, dtype: float64 |           Name: V248, dtype: float64 |
| Name: V248, dtype: float64 |           |
+-----+-----+-----+
```

The unique values in V248 feature is: 24
V248 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V249 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.042851	mean	1.025316	
mean	1.249706			
std	0.528812	std	0.494074	
std	0.807767			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	22.000000	max	22.000000	
max	18.000000			
Name: V249, dtype: float64		Name: V249, dtype: float64		
Name: V249, dtype: float64				

The unique values in V249 feature is: 24
V249 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V250 feature

Five Number Summary

```
+-----+-----+-----+
```

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	141416.000000	count	130434.000000
mean	0.788588	mean	0.782181
std	0.488386	std	0.442629
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	18.000000	max	17.000000
	Name: V250, dtype: float64		Name: V250, dtype: float64
	Name: V250, dtype: float64		

The unique values in V250 feature is: 20

V250 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V251 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	141416.000000	count	130434.000000
mean	0.794231	mean	0.785018
std	0.501240	std	0.447471
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	18.000000	max	17.000000
	Name: V251, dtype: float64		Name: V251, dtype: float64
	Name: V251, dtype: float64		

The unique values in V251 feature is: 20
V251 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V252 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	130430.000000	count	120238.000000
mean	1.032048	mean	1.014654
std	0.389546	std	0.269335
min	0.000000	min	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	24.000000	max	18.000000
Name: V252, dtype: float64		Name: V252, dtype: float64	
Name: V252, dtype: float64			

The unique values in V252 feature is: 26
V252 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V253 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions

```
+-----+-----+
|   count    130430.000000 |           count    120238.000000 |
count      10192.000000 |           mean     1.141494 |
|   mean      1.162164 |           mean     1.141494 |
mean      1.406005 |           std      3.727705 |
|   std      3.612553 |           std      3.727705 |
std      1.736321 |           min     0.000000 |
|   min      0.000000 |           min     0.000000 |
min      0.000000 |           25%     1.000000 |
|   25%     1.000000 |           25%     1.000000 |
25%     1.000000 |           50%     1.000000 |
|   50%     1.000000 |           50%     1.000000 |
50%     1.000000 |           75%     1.000000 |
|   75%     1.000000 |           75%     1.000000 |
75%     1.000000 |           max     163.000000 |
|   max     163.000000 |           max     163.000000 |
max     62.000000 |           Name: V253, dtype: float64 |
| Name: V253, dtype: float64 |           Name: V253, dtype: float64 |
Name: V253, dtype: float64 |
+-----+-----+
```

The unique values in V253 feature is: 67

V253 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V254 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count		count		count
count	130430.000000			120238.000000
mean	1.075113		mean	1.05539
mean	1.307790			
std	1.357550		std	1.36684
std	1.218921			
min	0.000000		min	0.00000
min	0.000000			
25%	1.000000		25%	1.00000
25%	1.000000			
50%	1.000000		50%	1.00000
50%	1.000000			
75%	1.000000		75%	1.00000
75%	1.000000			
max	60.000000		max	60.00000
max	27.000000			
Name: V254, dtype: float64		Name: V254, dtype: float64		
Name: V254, dtype: float64				

The unique values in V254 feature is: 46
V254 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V255 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	141416.000000	130434.000000	
mean	0.806656	0.796280	
std	0.929885	0.890288	
min	0.000000	0.000000	
25%	1.000000	1.000000	
50%	0.000000	1.000000	
75%	1.000000	1.000000	
max	87.000000	87.000000	
Name:	V255, dtype: float64	Name: V255, dtype: float64	
	Name: V255, dtype: float64		

The unique values in V255 feature is: 47
V255 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V256 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	141416.000000	130434.000000	
mean	0.806656	0.796280	
std	0.929885	0.890288	
min	0.000000	0.000000	
25%	1.000000	1.000000	
50%	0.000000	1.000000	
75%	1.000000	1.000000	
max	87.000000	87.000000	
Name:	V256, dtype: float64	Name: V256, dtype: float64	
	Name: V256, dtype: float64		

1. self_case_study_1_IEEE_fraud_detection_EDA

count	141416.000000	count	130434.000000
count	10982.000000	mean	0.800021
mean	0.814950	mean	0.800021
mean	0.992260	std	0.903193
std	0.969953	std	0.903193
std	1.546662	min	0.000000
min	0.000000	min	0.000000
min	0.000000	25%	1.000000
25%	0.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
75%	1.000000	max	87.000000
max	41.000000	max	87.000000
max	87.000000	Name: V256, dtype: float64	Name: V256, dtype: float64
Name: V256, dtype: float64			

The unique values in V256 feature is: 49

V256 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V257 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000	mean	1.106015	
mean	1.250993	mean	1.106015	
mean	2.961342	std	0.509331	
std	1.299956	std	0.509331	
std	3.923475	min	0.000000	
min	0.000000	min	0.000000	
min	0.000000	25%	1.000000	
25%	1.000000	25%	1.000000	
50%	1.000000	50%	1.000000	
50%	2.000000	50%	1.000000	
50%	2.000000	75%	1.000000	
75%	3.000000	75%	1.000000	
75%	3.000000	max	22.000000	
max	48.000000	max	22.000000	
max	48.000000	Name: V257, dtype: float64	Name: V257, dtype: float64	
Name: V257, dtype: float64				

The unique values in V257 feature is: 50
 V257 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V258 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	1.343510	mean	1.169115	
mean	3.400903			
std	2.015811	std	1.519058	
std	4.493237			
min	0.000000	min	0.000000	
min	0.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	2.000000			
75%	1.000000	75%	1.000000	
75%	4.000000			
max	66.000000	max	66.000000	
max	48.000000			
Name: V258, dtype: float64		Name: V258, dtype: float64		
Name: V258, dtype: float64				

The unique values in V258 feature is: 68
 V258 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V259 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	141416.000000	count	130434.000000	
count	10982.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	0.967832	mean	0.874841
mean	2.072300	std	1.884947
std	2.113447	std	3.741202
min	0.000000	min	0.000000
min	0.000000	25%	0.000000
25%	1.000000	25%	1.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
75%	2.000000	max	285.000000
max	285.000000	max	285.000000
max	71.000000		
	Name: V259, dtype: float64		Name: V259, dtype: float64
	Name: V259, dtype: float64		

The unique values in V259 feature is: 69

V259 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V260 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	0.964425	mean	0.950615	
mean	1.127355	std	0.338300	
std	0.379828	std	0.683625	
std	0.683625	min	0.000000	
min	0.000000	min	0.000000	
25%	1.000000	25%	1.000000	
25%	1.000000	50%	1.000000	
50%	1.000000	50%	1.000000	
75%	1.000000	75%	1.000000	
75%	1.000000	max	6.000000	
max	8.000000	max	8.000000	
max	8.000000			
	Name: V260, dtype: float64		Name: V260, dtype: float64	
	Name: V260, dtype: float64			

The unique values in V260 feature is: 10

V260 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V261 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	130430.000000	count	120238.000000
mean	1.107161	mean	1.066793
std	1.583399	std	1.247890
min	1.323230	min	1.946924
25%	0.000000	25%	0.000000
50%	1.000000	50%	0.000000
75%	1.000000	75%	1.000000
max	2.000000	max	1.000000
	Name: V261, dtype: float64	Name: V261, dtype: float64	
	Name: V261, dtype: float64		

The unique values in V261 feature is: 42

V261 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V262 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	130430.000000	count	120238.000000
mean	1.013279	mean	1.013279
std	1.336244	std	0.985903
min	1.323230	min	1.946924
25%	0.000000	25%	0.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	2.000000	max	1.000000
	Name: V262, dtype: float64	Name: V262, dtype: float64	
	Name: V262, dtype: float64		

1. self_case_study_1_IEEE_fraud_detection_EDA

std	0.625455	std	0.543699
std	1.185672	min	0.000000
min	0.000000	25%	1.000000
25%	1.000000	50%	1.000000
50%	1.000000	75%	1.000000
75%	1.000000	max	20.000000
max	16.000000	Name: V262, dtype: float64	Name: V262, dtype: float64
Name: V262, dtype: float64			

The unique values in V262 feature is: 22

V262 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V263 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	130430.000000	count	120238.000000
count	10192.000000	mean	111.837490
mean	117.390676	mean	182.903241
std	1294.851543	std	381.478067
std	381.478067	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	44.976200	75%	0.000000
75%	200.000000	max	153600.000000
max	153600.000000	Name: V263, dtype: float64	Name: V263, dtype: float64
Name: V263, dtype: float64			

The unique values in V263 feature is: 10423

V263 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V264 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	130430.000000	120238.000000	count
mean	201.657617	195.432737	mean
std	275.094347	2374.122496	std
min	0.000000	0.000000	min
25%	0.000000	0.000000	25%
50%	78.924500	0.000000	50%
75%	33.593498	19.817801	75%
max	153600.000000	153600.000000	max
count	10192.000000	Name: V264, dtype: float64	Name: V264, dtype: float64
mean	10035.000000		
std	1 Name: V264, dtype: float64		

The unique values in V264 feature is: 13359

V264 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V265 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	130430.000000	120238.000000	count
mean	153.520534	146.871712	mean
std	231.958628	1666.639314	std
min	1605.512276	459.758418	min
25%			25%
50%			50%
75%			75%
max			max
count	10192.000000	Name: V265, dtype: float64	Name: V265, dtype: float64
mean	10035.000000		
std	1 Name: V265, dtype: float64		

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
50%	62.353401		
75%	20.897525	75%	0.000000
75%	256.954865		
max	153600.000000	max	153600.000000
max	5475.000000		
Name: V265, dtype: float64		Name: V265, dtype: float64	
Name: V265, dtype: float64			

The unique values in V265 feature is: 11758

V265 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V266 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	9.167839	mean	9.486243	
mean	5.411525			
std	208.038863	std	216.513053	
std	28.666797			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	635.000000			
Name: V266, dtype: float64		Name: V266, dtype: float64		
Name: V266, dtype: float64				

The unique values in V266 feature is: 2179

V266 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V267 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	36.525103	mean	37.912557	
mean	20.156896			
std	644.878586	std	671.228294	
std	80.393649			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	1495.000000			
Name: V267, dtype: float64		Name: V267, dtype: float64		
Name: V267, dtype: float64				

The unique values in V267 feature is: 3617

V267 feature has 460110 null values, which holds 0.7791343516103905 %

```
*****
*****
```

Analysis of V268 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	18.813407	mean	19.281140	
mean	13.295416			
std	311.280181	std	323.675597	
std	63.343620			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
75%	0.000000	max	55125.000000
max	1495.000000	max	55125.000000
Name: V268, dtype: float64		Name: V268, dtype: float64	
Name: V268, dtype: float64			

The unique values in V268 feature is: 2757

V268 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V269 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	5.997508	mean	6.353898	
mean	1.793073			
std	207.317539	std	215.823094	
std	22.423101			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	635.000000			
Name: V269, dtype: float64		Name: V269, dtype: float64		
Name: V269, dtype: float64				

The unique values in V269 feature is: 152

V269 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V270 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	141416.000000	count	130434.000000	
count	10982.000000			
mean	7.712764	mean	7.414661	
mean	11.253346			
std	65.507232	std	66.759462	
std	48.069232			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	4000.000000	max	4000.000000	
max	800.000000			
Name: V270, dtype: float64		Name: V270, dtype: float64		
Name: V270, dtype: float64				

The unique values in V270 feature is: 2341

V270 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V271 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	141416.000000	count	130434.000000	
count	10982.000000			
mean	9.445133	mean	8.858882	
mean	16.408079			
std	74.147726	std	74.100747	
std	74.354161			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	0.000000		50%	0.000000	
50%		0.000000				
	75%	0.000000		75%	0.000000	
75%		0.000000				
	max	4000.000000		max	4000.000000	
max		1495.000000				
	Name: V271, dtype: float64			Name: V271, dtype: float64		
Name: V271, dtype: float64						
<hr/>						
<hr/>						

The unique values in V271 feature is: 2788

V271 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V272 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	141416.000000		count	130434.000000	
count	10982.000000				
mean	8.464571		mean	7.951063	
mean	14.563545				
std	69.723735		std	69.840441	
std	68.029692				
min	0.000000		min	0.000000	
min	0.000000				
25%	0.000000		25%	0.000000	
25%	0.000000				
50%	0.000000		50%	0.000000	
50%	0.000000				
75%	0.000000		75%	0.000000	
75%	0.000000				
max	4000.000000		max	4000.000000	
max	1495.000000				
	Name: V272, dtype: float64			Name: V272, dtype: float64	
Name: V272, dtype: float64					
<hr/>					
<hr/>					

The unique values in V272 feature is: 2508

V272 feature has 449124 null values, which holds 0.7605310393876791 %

Analysis of V273 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	73.825549	mean	69.557475	
mean	124.177259			
std	935.187927	std	969.914136	
std	302.272394			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	121.081802			
max	51200.000000	max	51200.000000	
max	5450.000000			
Name: V273, dtype: float64		Name: V273, dtype: float64		
Name: V273, dtype: float64				

The unique values in V273 feature is: 7178

V273 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V274 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	107.151636	mean	101.464798	
mean	174.240926			
std	1258.734139	std	1306.320800	
std	373.562643			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	20.558701			

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	0.000000	75%	0.000000
75%	179.749496		
max	66000.000000	max	66000.000000
max	5450.000000		
Name: V274, dtype: float64		Name: V274, dtype: float64	
Name: V274, dtype: float64			

The unique values in V274 feature is: 8316
V274 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V275 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	130430.000000	count	120238.000000
count	10192.000000		
mean	88.899939	mean	83.552263
mean	151.988036		
std	1072.883139	std	1112.893253
std	339.219874		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	10.000000		
75%	0.000000	75%	0.000000
75%	153.360271		
max	51200.000000	max	51200.000000
max	5450.000000		
Name: V275, dtype: float64		Name: V275, dtype: float64	
Name: V275, dtype: float64			

The unique values in V275 feature is: 7777
V275 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V276 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	31.797277	mean	30.062054	
mean	52.268202			
std	615.659714	std	638.347369	
std	207.245465			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	104060.000000	max	104060.000000	
max	4500.000000			
Name: V276, dtype: float64		Name: V276, dtype: float64		
Name: V276, dtype: float64				

The unique values in V276 feature is: 2264

V276 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V277 feature***Five Number Summary***

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	51.956645	mean	49.667822	
mean	78.958558			
std	732.145368	std	757.231698	
std	307.370447			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	17.106775			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	104060.000000	max	104060.000000
max	6390.000000		
Name: V277, dtype: float64		Name: V277, dtype: float64	
Name: V277, dtype: float64			

The unique values in V277 feature is: 2541
V277 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V278 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	130430.000000	count	120238.000000	
count	10192.000000			
mean	42.328228	mean	40.397293	
mean	65.108035			
std	660.611820	std	684.217058	
std	247.687530			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	0.000000	75%	0.000000	
75%	0.000000			
max	104060.000000	max	104060.000000	
max	4500.000000			
Name: V278, dtype: float64		Name: V278, dtype: float64		
Name: V278, dtype: float64				

The unique values in V278 feature is: 2399
V278 feature has 460110 null values, which holds 0.7791343516103905 %

Analysis of V279 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	1.123061	mean	1.138076	
mean	0.708920			
std	21.021950	std	21.396452	
std	1.894142			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	880.000000	max	880.000000	
max	55.000000			
Name: V279, dtype: float64		Name: V279, dtype: float64		
Name: V279, dtype: float64				

The unique values in V279 feature is: 882

V279 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V280 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	1.967082	mean	1.971230	
mean	1.852669			
std	27.851780	std	28.302649	
std	8.797109			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	1.000000	75%	1.000000	
75%	2.000000			
max	975.000000	max	975.000000	
max	243.000000			

```
| Name: V280, dtype: float64 |           Name: V280, dtype: float64 |
| Name: V280, dtype: float64 |           |
+-----+-----+-----+
```

The unique values in V280 feature is: 976
V280 feature has 12 null values, which holds 2.0320384732617603e-05 %

```
*****
*****
```

Analysis of V281 feature

Five Number Summary

```
+-----+-----+-----+
|-----+-----+-----+
| Feature Description      | Feature Description of legit transactions | Feature
| Description of fraud transactions |                               |
+-----+-----+-----+
-----+
| count      589271.000000 |           count      568654.000000 |
count      20617.000000 |           |           mean      0.080692 |
| mean       0.087783 |           |           std       0.484742 |
mean       0.283358 |           |           min       0.000000 |
| std        0.512748 |           |           |           0.000000 |
std        0.996906 |           |           |           0.000000 |
| min        0.000000 |           |           |           0.000000 |
min        0.000000 |           |           |           0.000000 |
| 25%        0.000000 |           |           25%       0.000000 |
25%        0.000000 |           |           |           0.000000 |
| 50%        0.000000 |           |           50%       0.000000 |
50%        0.000000 |           |           |           0.000000 |
| 75%        0.000000 |           |           75%       0.000000 |
75%        0.000000 |           |           |           0.000000 |
| max        22.000000 |           |           max       22.000000 |
max        15.000000 |           |           |           22.000000 |
| Name: V281, dtype: float64 |           Name: V281, dtype: float64 |
Name: V281, dtype: float64 |           |
+-----+-----+-----+
-----+
```

The unique values in V281 feature is: 24
V281 feature has 1269 null values, which holds 0.0021488806854743116 %

```
*****
*****
```

Analysis of V282 feature

Five Number Summary

```
+-----+-----+-----+
-----+
```

1. self_case_study_1_IEEE_fraud_detection_EDA

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	589271.000000	count	568654.000000
mean	0.817171	mean	0.798894
std	0.921880	std	0.898346
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	32.000000	max	32.000000
	Name: V282, dtype: float64		Name: V282, dtype: float64
	Name: V282, dtype: float64		

The unique values in V282 feature is: 34

V282 feature has 1269 null values, which holds 0.0021488806854743116 %

Analysis of V283 feature

Five Number Summary

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	589271.000000	count	568654.000000
mean	0.991114	mean	0.957656
std	1.558731	std	1.474404
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	1.000000	50%	1.000000
75%	1.000000	75%	1.000000
max	68.000000	max	68.000000
	Name: V283, dtype: float64		Name: V283, dtype: float64
	Name: V283, dtype: float64		

```
+-----+-----+
-----+-----+
```

The unique values in V283 feature is: 63
V283 feature has 1269 null values, which holds 0.0021488806854743116 %

```
*****  
*****
```

Analysis of V284 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+-----+
| count      590528.000000 | | count      569867.000000 |
count      20661.000000 | | mean       0.086267 |
mean       0.151300 | | std        0.334147 |
std        0.435648 | | min        0.000000 |
min        0.000000 | | max        12.000000 |
max        6.000000 | | Name: V284, dtype: float64 |
Name: V284, dtype: float64 | | Name: V284, dtype: float64 |
+-----+-----+-----+
```

The unique values in V284 feature is: 14
V284 feature has 12 null values, which holds 2.0320384732617603e-05 %

```
*****  
*****
```

Analysis of V285 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
```

1. self case study 1 IEEE fraud detection EDA

count	590528.000000		count	569867.000000	
count	20661.000000				
mean	1.167660		mean	1.173365	
mean	1.010309				
std	3.282454		std	3.314839	
std	2.203654				
min	0.000000		min	0.000000	
min	0.000000				
25%	0.000000		25%	0.000000	
25%	0.000000				
50%	0.000000		50%	0.000000	
50%	0.000000				
75%	1.000000		75%	1.000000	
75%	1.000000				
max	95.000000		max	95.000000	
max	22.000000				
Name: V285, dtype: float64			Name: V285, dtype: float64		
Name: V285, dtype: float64					

The unique values in V285 feature is: 97

V285 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V286 feature

Five Number Summary

Feature Description of fraud transactions		Feature Description of legit transactions		Feature
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	0.031492	mean	0.031332	
mean	0.035913			
std	0.190939	std	0.190495	
std	0.202758			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	8.000000	max	8.000000	
max	3.000000			
Name: V286, dtype: float64		Name: V286, dtype: float64		
Name: V286, dtype: float64				

The unique values in V286 feature is: 10
 V286 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V287 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590528.000000	569867.000000	
mean	0.358579	0.354384	
std	0.474275	1.081547	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	31.000000	31.000000	
	Name: V287, dtype: float64	Name: V287, dtype: float64	
	Name: V287, dtype: float64		

The unique values in V287 feature is: 33
 V287 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V288 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590528.000000	569867.000000	
mean	0.358579	0.354384	
std	0.474275	1.081547	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	31.000000	31.000000	
	Name: V288, dtype: float64	Name: V288, dtype: float64	
	Name: V288, dtype: float64		

1. self_case_study_1_IEEE_fraud_detection_EDA

count 589271.000000	count 568654.000000
count 20617.000000	
mean 0.184350	mean 0.181451
mean 0.264296	
std 0.430989	std 0.426852
std 0.526318	
min 0.000000	min 0.000000
min 0.000000	
25% 0.000000	25% 0.000000
25% 0.000000	
50% 0.000000	50% 0.000000
50% 0.000000	
75% 0.000000	75% 0.000000
75% 0.000000	
max 10.000000	max 10.000000
max 4.000000	
Name: V288, dtype: float64	Name: V288, dtype: float64
Name: V288, dtype: float64	

The unique values in V288 feature is: 12

V288 feature has 1269 null values, which holds 0.0021488806854743116 %

Analysis of V289 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count 589271.000000	count 568654.000000	count 568654.000000	count 568654.000000	
count 20617.000000		mean 0.231111	mean 0.231111	
mean 0.235975	mean 0.231111	mean 0.231111	mean 0.231111	
mean 0.370131		std 0.589855	std 0.589855	
std 0.599231	std 0.589855	std 0.589855	std 0.589855	
std 0.804965		min 0.000000	min 0.000000	
min 0.000000	min 0.000000	min 0.000000	min 0.000000	
min 0.000000		25% 0.000000	25% 0.000000	
25% 0.000000	25% 0.000000	25% 0.000000	25% 0.000000	
25% 0.000000				
50% 0.000000	50% 0.000000	50% 0.000000	50% 0.000000	
50% 0.000000				
75% 0.000000	75% 0.000000	75% 0.000000	75% 0.000000	
75% 0.000000				
max 12.000000	max 12.000000	max 12.000000	max 12.000000	
max 7.000000				
Name: V289, dtype: float64	Name: V289, dtype: float64	Name: V289, dtype: float64	Name: V289, dtype: float64	
Name: V289, dtype: float64				

The unique values in V289 feature is: 14
 V289 feature has 1269 null values, which holds 0.0021488806854743116 %

Analysis of V290 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	1.103011	mean	1.097133	
mean	1.265137			
std	0.768897	std	0.732916	
std	1.433346			
min	1.000000	min	1.000000	
min	1.000000			
25%	1.000000	25%	1.000000	
25%	1.000000			
50%	1.000000	50%	1.000000	
50%	1.000000			
75%	1.000000	75%	1.000000	
75%	1.000000			
max	67.000000	max	67.000000	
max	56.000000			
Name: V290, dtype: float64		Name: V290, dtype: float64		
Name: V290, dtype: float64				

The unique values in V290 feature is: 59
 V290 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V291 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	1.659811	mean	1.596555
mean	3.40453	std	14.822321
std	16.252538	min	1.000000
std	38.56033	25%	1.000000
min	1.000000	50%	1.000000
25%	1.000000	75%	1.000000
50%	1.000000	max	1055.000000
75%	1.000000		1055.000000
max	2.00000		
	1055.000000	Name: V291, dtype: float64	Name: V291, dtype: float64
	1052.000000		
		+-----+	+-----+
		-+-----+	-+-----+

The unique values in V291 feature is: 220

V291 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V292 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	1.239916	mean	1.220667	
mean	1.770824	std	3.512115	
std	3.775050			
std	8.173748			
min	1.000000	min	1.000000	
min	1.000000	25%	1.000000	
25%	1.000000	50%	1.000000	
50%	1.000000	75%	1.000000	
75%	1.000000	max	323.000000	
max	323.000000			
	244.000000			
		Name: V292, dtype: float64	Name: V292, dtype: float64	
		+-----+	+-----+	+-----+
		-+-----+	-+-----+	-+-----+

The unique values in V292 feature is: 174

V292 feature has 12 null values, which holds 2.0320384732617603e-05 %

```
*****
*****
```

Analysis of V293 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	0.942599	mean	0.960243	
mean	0.455931			
std	20.588816	std	20.955974	
std	1.713467			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	869.000000	max	869.000000	
max	55.000000			
Name: V293, dtype: float64		Name: V293, dtype: float64		
Name: V293, dtype: float64				

The unique values in V293 feature is: 871

V293 feature has 12 null values, which holds 2.0320384732617603e-05 %

```
*****
*****
```

Analysis of V294 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	2.313863	mean	2.281948	
mean	3.194134			

1. self_case_study_1_IEEE_fraud_detection_EDA

std	39.526468	std	39.557319
std	38.656394	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	1.000000	max	1286.000000
max	1078.000000	Name: V294, dtype: float64	Name: V294, dtype: float64
Name: V294, dtype: float64			
<hr/>			
<hr/>			

The unique values in V294 feature is: 1287

V294 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V295 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	590528.000000	count	569867.000000
count	20661.000000	mean	1.441736
mean	1.204153	std	26.379483
std	25.962948	min	0.000000
min	8.524928	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	75%	0.000000
75%	1.000000	max	928.000000
max	243.000000	Name: V295, dtype: float64	Name: V295, dtype: float64
Name: V295, dtype: float64			
<hr/>			
<hr/>			

The unique values in V295 feature is: 929

V295 feature has 12 null values, which holds 2.0320384732617603e-05 %

```
*****
*****
```

Analysis of V296 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	589271.000000	568654.000000	
mean	0.328917	0.326330	
std	3.264745	3.289619	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	93.000000	93.000000	
Name: V296, dtype: float64		Name: V296, dtype: float64	
Name: V296, dtype: float64			

The unique values in V296 feature is: 95

V296 feature has 1269 null values, which holds 0.0021488806854743116 %

```
*****
*****
```

Analysis of V297 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590528.000000	569867.000000	
mean	0.089034	0.088731	
std	0.097382	0.632400	
std	0.628352	0.504005	

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000	min	0.000000
min	0.000000	25%	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	12.000000	max	12.000000
max	12.000000		
Name: V297, dtype: float64		Name: V297, dtype: float64	
Name: V297, dtype: float64			

The unique values in V297 feature is: 14

V297 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V298 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	0.298829	mean	0.298047	
mean	0.320410			
std	3.175062	std	3.212883	
std	1.848509			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	93.000000	max	93.000000	
max	47.000000			
Name: V298, dtype: float64		Name: V298, dtype: float64		
Name: V298, dtype: float64				

The unique values in V298 feature is: 95

V298 feature has 12 null values, which holds 2.0320384732617603e-05 %

```
*****
*****
```

Analysis of V299 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	0.171655	mean	0.171756	
mean	0.168869			
std	1.724218	std	1.747153	
std	0.881503			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	49.000000	max	49.000000	
max	23.000000			
Name: V299, dtype: float64		Name: V299, dtype: float64		
Name: V299, dtype: float64				

The unique values in V299 feature is: 51

V299 feature has 12 null values, which holds 2.0320384732617603e-05 %

```
*****
*****
```

Analysis of V300 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	589271.000000	count	568654.000000	
count	20617.000000			
mean	0.045507	mean	0.044111	
mean	0.084008			
std	0.289573	std	0.283579	
std	0.420798			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	0.000000	25%	0.000000
25%	0.000000	50%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
75%	0.000000	max	11.000000
max	11.000000	max	11.000000
Name: V300, dtype: float64		Name: V300, dtype: float64	
Name: V300, dtype: float64			

The unique values in V300 feature is: 13

V300 feature has 1269 null values, which holds 0.0021488806854743116 %

Analysis of V301 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	589271.000000	count	568654.000000	
count	20617.000000			
mean	0.052002	mean	0.050057	
mean	0.105641			
std	0.318310	std	0.309888	
std	0.494239			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	13.000000	max	13.000000	
max	11.000000			
Name: V301, dtype: float64		Name: V301, dtype: float64		
Name: V301, dtype: float64				

The unique values in V301 feature is: 15

V301 feature has 1269 null values, which holds 0.0021488806854743116 %

Analysis of V302 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	0.251761	mean	0.239566	
mean	0.588113			
std	0.481889	std	0.472353	
std	0.604951			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	1.000000	50%	0.000000	
50%	1.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	16.000000	max	16.000000	
max	7.000000			
Name: V302, dtype: float64		Name: V302, dtype: float64		
Name: V302, dtype: float64				

The unique values in V302 feature is: 18

V302 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V303 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	0.283140	mean	0.268382	
mean	0.690189			
std	0.623608	std	0.608920	
std	0.846492			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	0.000000		50%	0.000000	
50%		1.000000				
	75%	0.000000		75%	0.000000	
75%		1.000000				
	max	20.000000		max	20.000000	
max		15.000000				
	Name: V303, dtype: float64			Name: V303, dtype: float64		
Name: V303, dtype: float64						
<hr/>						
<hr/>						

The unique values in V303 feature is: 22

V303 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V304 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000		count	569867.000000
count	20661.000000			
mean	0.264208		mean	0.250981
mean	0.629011			
std	0.528238		std	0.516766
std	0.686876			
min	0.000000		min	0.000000
min	0.000000			
25%	0.000000		25%	0.000000
25%	0.000000			
50%	0.000000		50%	0.000000
50%	1.000000			
75%	0.000000		75%	0.000000
75%	1.000000			
max	16.000000		max	16.000000
max	9.000000			
	Name: V304, dtype: float64		Name: V304, dtype: float64	
Name: V304, dtype: float64				
<hr/>				
<hr/>				

The unique values in V304 feature is: 18

V304 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V305 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	590528.000000	569867.000000	
count	20661.0		
mean	1.000007	1.000007	
mean	1.0		
std	0.002603	0.002649	
std	0.0		
min	1.000000	1.000000	
min	1.0		
25%	1.000000	1.000000	
25%	1.0		
50%	1.000000	1.000000	
50%	1.0		
75%	1.000000	1.000000	
75%	1.0		
max	2.000000	2.000000	
max	1.0		
	Name: V305, dtype: float64	Name: V305, dtype: float64	
	Name: V305, dtype: float64		

The unique values in V305 feature is: 3

V305 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V306 feature

Five Number Summary

	Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	590528.000000	569867.000000	
count	20661.000000		
mean	139.748713	140.581994	
mean	116.765349		
std	2348.849634	2389.815422	
std	402.869710		
min	0.000000	0.000000	
min	0.000000		
25%	0.000000	0.000000	
25%	0.000000		
50%	0.000000	0.000000	
50%	0.000000		

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	0.000000	75%	0.000000
75%	54.225399		
max	108800.000000	max	108800.000000
	11848.000000		
Name: V306, dtype: float64		Name: V306, dtype: float64	
Name: V306, dtype: float64			

The unique values in V306 feature is: 16211
V306 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V307 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions
count	569867.000000	
count	20661.000000	
mean	403.708103	
mean	545.881593	
std	4431.742064	
std	4391.992977	
min	0.000000	
min	0.000000	
25%	0.000000	
25%	0.000000	
50%	0.000000	
50%	12.809000	
75%	0.000000	
75%	150.000000	
max	145765.000000	
max	273.850006	
	83258.367188	
Name: V307, dtype: float64		Name: V307, dtype: float64
Name: V307, dtype: float64		

The unique values in V307 feature is: 37368
V307 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V308 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	230.413180	mean	228.541233	
mean	282.044788			
std	3021.924247	std	3069.911131	
std	1032.666247			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	35.970001	75%	34.000000	
75%	156.774002			
max	108800.000000	max	108800.000000	
max	18123.957031			
Name: V308, dtype: float64		Name: V308, dtype: float64		
Name: V308, dtype: float64				

The unique values in V308 feature is: 23065

V308 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V309 feature***Five Number Summary***

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	10.995986	mean	10.547606	
mean	23.363125			
std	116.254277	std	116.620286	
std	104.910410			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

max	55125.000000	max	55125.000000
max	2793.000000		
	Name: V309, dtype: float64		Name: V309, dtype: float64
	Name: V309, dtype: float64		

The unique values in V309 feature is: 4237
V309 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V310 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	118.195658	mean	117.451559	
mean	138.719226			
std	352.983093	std	351.042468	
std	402.308240			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
50%	0.000000	50%	0.000000	
75%	107.949997	75%	107.949997	
75%	97.000000			
max	55125.000000	max	55125.000000	
max	5216.000000			
	Name: V310, dtype: float64		Name: V310, dtype: float64	
	Name: V310, dtype: float64			

The unique values in V310 feature is: 19137
V310 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V311 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions		
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	4.202175	mean	4.176831	
mean	4.901211			
std	102.374938	std	103.901884	
std	42.340216			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	1448.500000			
Name: V311, dtype: float64		Name: V311, dtype: float64		
Name: V311, dtype: float64				

The unique values in V311 feature is: 3099

V311 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V312 feature

Five Number Summary

Feature Description		Feature Description of legit transactions Feature Description of fraud transactions		
count	590528.000000	count	569867.000000	
count	20661.000000			
mean	39.173910	mean	37.942292	
mean	73.144119			
std	172.128339	std	169.942491	
std	221.490399			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	30.950001	75%	0.000000	
75%	30.950001			
max	55125.000000	max	55125.000000	
max	3981.000000			

Name: V312, dtype: float64	Name: V312, dtype: float64
Name: V312, dtype: float64	
+-----+-----+	+-----+-----+
-----+-----+	

The unique values in V312 feature is: 8069
V312 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V313 feature

Five Number Summary

Feature Description		Feature Description of legit transactions	Feature Description of fraud transactions
count	589271.000000	count	568654.000000
count	20617.000000		
mean	21.351473	mean	20.593754
mean	42.250735		
std	95.902970	std	93.852655
std	139.558705		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	0.000000		
max	4817.470215	max	4817.470215
max	2652.000000		
Name: V313, dtype: float64	Name: V313, dtype: float64		
Name: V313, dtype: float64			
+-----+-----+	+-----+-----+		
-----+-----+			

The unique values in V313 feature is: 5530
V313 feature has 1269 null values, which holds 0.0021488806854743116 %

Analysis of V314 feature

Five Number Summary

+-----+-----+	+-----+-----+
-----+-----+	

1. self_case_study_1_IEEE_fraud_detection_EDA

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	589271.000000	count	568654.000000
mean	43.319174	mean	42.045247
std	173.619028	std	169.612774
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	7519.870117	max	7519.870117
max	3981.000000		
	Name: V314, dtype: float64	Name: V314, dtype: float64	
	Name: V314, dtype: float64		

The unique values in V314 feature is: 11378

V314 feature has 1269 null values, which holds 0.0021488806854743116 %

Analysis of V315 feature

Five Number Summary

	Feature Description Description of fraud transactions	Feature Description of legit transactions	Feature
count	589271.000000	count	568654.000000
mean	26.806977	mean	25.732347
std	116.853222	std	113.346437
min	0.000000	min	0.000000
25%	0.000000	25%	0.000000
50%	0.000000	50%	0.000000
75%	0.000000	75%	0.000000
max	4817.470215	max	4817.470215
max	3981.000000		
	Name: V315, dtype: float64	Name: V315, dtype: float64	
	Name: V315, dtype: float64		

```
+-----+-----+
-----+-----+
```

The unique values in V315 feature is: 6974
V315 feature has 1269 null values, which holds 0.0021488806854743116 %

```
*****  
*****
```

Analysis of V316 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
+-----+-----+-----+
| count      590528.000000 | | count      569867.000000 |
count      20661.000000 | | mean      111.098084 |
mean      74.526581 | | std       2309.921526 |
std       336.081637 | | min       0.000000 |
min       0.000000 | | 25%      0.000000 |
25%      0.000000 | | 50%      0.000000 |
50%      0.000000 | | 75%      0.000000 |
75%      0.000000 | | max       93736.000000 |
max      11848.000000 | | Name: V316, dtype: float64 | | Name: V316, dtype: float64 |
Name: V316, dtype: float64 | +-----+-----+-----+
```

The unique values in V316 feature is: 9815
V316 feature has 12 null values, which holds 2.0320384732617603e-05 %

```
*****  
*****
```

Analysis of V317 feature

Five Number Summary

```
+-----+-----+
-----+-----+
| Feature Description | Feature Description of legit transactions | Feature
Description of fraud transactions |
```

count	590528.000000	count	569867.000000
count	20661.000000	mean	243.810156
mean	247.606741	std	4013.364529
mean	352.323304	min	0.000000
std	3980.042828	25%	0.000000
std	2912.608406	50%	0.000000
min	0.000000	75%	0.000000
min	0.000000	max	134021.000000
25%	0.000000	max	134021.000000
50%	0.000000	max	134021.000000
75%	0.000000	max	134021.000000
75%	65.004501	max	134021.000000
max	134021.000000	Name: V317, dtype: float64	Name: V317, dtype: float64
max	82130.953125	Name: V317, dtype: float64	Name: V317, dtype: float64

The unique values in V317 feature is: 15185
V317 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V318 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	590528.000000	count	569867.000000	
count	20661.000000	mean	161.623043	
mean	162.153398	std	2838.427857	
mean	176.781527	min	0.000000	
std	2793.343636	25%	0.000000	
std	894.133506	50%	0.000000	
min	0.000000	75%	0.000000	
min	0.000000	max	98476.000000	
25%	0.000000	max	98476.000000	
25%	0.000000	max	98476.000000	
50%	0.000000	max	98476.000000	
50%	0.000000	max	98476.000000	
75%	0.000000	max	98476.000000	
75%	28.397400	max	98476.000000	
max	98476.000000	Name: V318, dtype: float64	Name: V318, dtype: float64	
max	18123.957031	Name: V318, dtype: float64	Name: V318, dtype: float64	

The unique values in V318 feature is: 12310
 V318 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V319 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590528.000000	569867.000000	
mean	18.372476	18.368631	
std	18.478514	337.454879	
min	332.304848	123.652990	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	4528.899902	104060.000000	
Name:	V319, dtype: float64	Name: V319, dtype: float64	
	Name: V319, dtype: float64		

The unique values in V319 feature is: 4800
 V319 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V320 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	590528.000000	569867.000000	
mean	18.372476	18.368631	
std	18.478514	337.454879	
min	332.304848	123.652990	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	4528.899902	104060.000000	
Name:	V319, dtype: float64	Name: V319, dtype: float64	
	Name: V319, dtype: float64		

1. self_case_study_1_IEEE_fraud_detection_EDA

count 590528.000000	count 569867.000000
count 20661.000000	
mean 42.073133	mean 41.625866
mean 54.409545	
std 473.499307	std 478.434705
std 307.351165	
min 0.000000	min 0.000000
min 0.000000	
25% 0.000000	25% 0.000000
25% 0.000000	
50% 0.000000	50% 0.000000
50% 0.000000	
75% 0.000000	75% 0.000000
75% 0.000000	
max 104060.000000	max 104060.000000
max 5794.500000	
Name: V320, dtype: float64	Name: V320, dtype: float64
Name: V320, dtype: float64	

The unique values in V320 feature is: 6440

V320 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V321 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count 590528.000000	count 569867.000000	count 569867.000000	count 569867.000000	
count 20661.000000		mean 28.204580	mean 28.204580	
mean 28.326584		mean 28.204580	mean 28.204580	
mean 31.691673		std 387.440031	std 387.440031	
std 382.053171		std 387.440031	std 387.440031	
std 177.831364		min 0.000000	min 0.000000	
min 0.000000		min 0.000000	min 0.000000	
min 0.000000		25% 0.000000	25% 0.000000	
25% 0.000000		25% 0.000000	25% 0.000000	
25% 0.000000		50% 0.000000	50% 0.000000	
50% 0.000000		50% 0.000000	50% 0.000000	
50% 0.000000		75% 0.000000	75% 0.000000	
75% 0.000000		75% 0.000000	75% 0.000000	
75% 0.000000		max 104060.000000	max 104060.000000	
max 104060.000000		max 104060.000000	max 104060.000000	
max 4528.899902				
Name: V321, dtype: float64		Name: V321, dtype: float64		
Name: V321, dtype: float64				

The unique values in V321 feature is: 5561
 V321 feature has 12 null values, which holds 2.0320384732617603e-05 %

Analysis of V322 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	6.220289	mean	6.481510	
mean	0.645415			
std	56.022561	std	57.306006	
std	1.371503			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	880.000000	max	880.000000	
max	16.000000			
Name: V322, dtype: float64		Name: V322, dtype: float64		
Name: V322, dtype: float64				

The unique values in V322 feature is: 882
 V322 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V323 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

mean	13.103775	mean	13.642789
mean	1.600380		
std	106.739813	std	109.179292
std	3.709768		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	1.000000	75%	1.000000
75%	1.000000		
max	1411.000000	max	1411.000000
max	34.000000		
Name: V323, dtype: float64		Name: V323, dtype: float64	
Name: V323, dtype: float64			

The unique values in V323 feature is: 1412

V323 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V324 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	9.184612	mean	9.567215	
mean	1.019262			
std	73.627893	std	75.309861	
std	2.257605			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	1.000000			
max	976.000000	max	976.000000	
max	24.000000			
Name: V324, dtype: float64		Name: V324, dtype: float64		
Name: V324, dtype: float64				

The unique values in V324 feature is: 977

V324 feature has 508189 null values, which holds 0.8605496664070174 %

```
*****
*****
```

Analysis of V325 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	82351.000000	count	78665.000000
count	3686.000000		
mean	0.058494	mean	0.057980
mean	0.069452		
std	0.304415	std	0.301833
std	0.354930		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	0.000000		
max	12.000000	max	12.000000
max	6.000000		
Name: V325, dtype: float64		Name: V325, dtype: float64	
Name: V325, dtype: float64			

The unique values in V325 feature is: 14

V325 feature has 508189 null values, which holds 0.8605496664070174 %

```
*****
*****
```

Analysis of V326 feature

Five Number Summary

	Feature Description Discription of fraud transactions	Feature Description of legit transactions	Feature Discription of legit transactions
count	82351.000000	count	78665.000000
count	3686.000000		
mean	0.851040	mean	0.856887
mean	0.726262		

1. self_case_study_1_IEEE_fraud_detection_EDA

std	3.950295	std	3.996386
std	2.788170		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	0.000000		
max	44.000000	max	44.000000
max	27.000000		
Name: V326, dtype: float64		Name: V326, dtype: float64	
Name: V326, dtype: float64			

The unique values in V326 feature is: 46

V326 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V327 feature

Five Number Summary

Feature Description	Feature Description of legit transactions	Feature Description of fraud transactions	
count	82351.000000	count	78665.000000
count	3686.000000		
mean	0.296633	mean	0.296943
mean	0.290016		
std	1.364356	std	1.369157
std	1.257701		
min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	0.000000		
max	18.000000	max	14.000000
max	18.000000		
Name: V327, dtype: float64		Name: V327, dtype: float64	
Name: V327, dtype: float64			

The unique values in V327 feature is: 20

V327 feature has 508189 null values, which holds 0.8605496664070174 %

```
*****
*****
```

Analysis of V328 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	82351.000000	78665.000000	
count	3686.000000		
mean	0.336790	0.340571	
mean	0.256104		
std	1.580144	1.607652	
std	0.786631		
min	0.000000	0.000000	
min	0.000000		
25%	0.000000	0.000000	
25%	0.000000		
50%	0.000000	0.000000	
50%	0.000000		
75%	0.000000	0.000000	
75%	0.000000		
max	15.000000	15.000000	
max	11.000000		
Name: V328, dtype: float64		Name: V328, dtype: float64	
Name: V328, dtype: float64			

The unique values in V328 feature is: 17

V328 feature has 508189 null values, which holds 0.8605496664070174 %

```
*****
*****
```

Analysis of V329 feature

Five Number Summary

	Feature Description of fraud transactions	Feature Description of legit transactions	Feature Description
count	82351.000000	78665.000000	
count	3686.000000		
mean	1.312844	1.356690	
mean	0.377103		
std	8.769083	8.966832	
std	1.063497		

1. self_case_study_1_IEEE_fraud_detection_EDA

min	0.000000	min	0.000000
min	0.000000		
25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	0.000000	75%	0.000000
75%	0.000000		
max	99.000000	max	99.000000
max	13.000000		
Name: V329, dtype: float64		Name: V329, dtype: float64	
Name: V329, dtype: float64			

The unique values in V329 feature is: 101

V329 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V330 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	0.775874	mean	0.797534	
mean	0.313619			
std	4.727971	std	4.831921	
std	0.960475			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55.000000	max	55.000000	
max	13.000000			
Name: V330, dtype: float64		Name: V330, dtype: float64		
Name: V330, dtype: float64				

The unique values in V330 feature is: 57

V330 feature has 508189 null values, which holds 0.8605496664070174 %

```
*****
*****
```

Analysis of V331 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	721.741883	mean	751.325339	
mean	90.384699			
std	6217.223583	std	6359.489982	
std	227.643019			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	75.000000			
max	160000.000000	max	160000.000000	
max	2600.000000			
Name: V331, dtype: float64		Name: V331, dtype: float64		
Name: V331, dtype: float64				

The unique values in V331 feature is: 1759

V331 feature has 508189 null values, which holds 0.8605496664070174 %

```
*****
*****
```

Analysis of V332 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	1375.783644	mean	1432.523827	
mean	164.859468			
std	11169.275702	std	11424.424611	
std	436.287131			
min	0.000000	min	0.000000	
min	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

25%	0.000000	25%	0.000000
25%	0.000000		
50%	0.000000	50%	0.000000
50%	0.000000		
75%	25.000000	75%	20.000000
75%	150.000000		
max	160000.000000	max	160000.000000
max	4700.000000		
Name: V332, dtype: float64		Name: V332, dtype: float64	
Name: V332, dtype: float64			

The unique values in V332 feature is: 2454

V332 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V333 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	1014.622782	mean	1056.184360	
mean	127.633749			
std	7955.735482	std	8137.291377	
std	339.856944			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	100.000000			
max	160000.000000	max	160000.000000	
max	4285.000000			
Name: V333, dtype: float64		Name: V333, dtype: float64		
Name: V333, dtype: float64				

The unique values in V333 feature is: 1972

V333 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V334 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	9.807015	mean	9.830808	
mean	9.299240			
std	243.861391	std	249.260209	
std	51.502023			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	55125.000000	max	55125.000000	
max	635.000000			
Name: V334, dtype: float64		Name: V334, dtype: float64		
Name: V334, dtype: float64				

The unique values in V334 feature is: 144

V334 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V335 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	59.16455	mean	59.622346	
mean	49.394466			
std	387.62948	std	394.187771	
std	201.848261			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

	50%	0.00000		50%	0.000000	
50%		0.000000				
	75%	0.00000		75%	0.000000	
75%		0.000000				
	max	55125.00000		max	55125.000000	
max		2430.000000				
	Name: V335, dtype: float64			Name: V335, dtype: float64		
Name: V335, dtype: float64						
	+-----+			+-----+		
	-----+			-----+		

The unique values in V335 feature is: 673

V335 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V336 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions	
count	82351.000000		count	78665.000000	
count	3686.000000				
mean	28.530903		mean	28.673672	
mean	25.483993				
std	274.576920		std	279.773375	
std	117.938993				
min	0.000000		min	0.000000	
min	0.000000				
25%	0.000000		25%	0.000000	
25%	0.000000				
50%	0.000000		50%	0.000000	
50%	0.000000				
75%	0.000000		75%	0.000000	
75%	0.000000				
max	55125.000000		max	55125.000000	
max	1605.000000				
	Name: V336, dtype: float64			Name: V336, dtype: float64	
Name: V336, dtype: float64					
	+-----+			+-----+	
	-----+			-----+	

The unique values in V336 feature is: 357

V336 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V337 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	55.352422	mean	56.177580	
mean	37.742268			
std	668.486833	std	683.262438	
std	142.505208			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			
75%	0.000000	75%	0.000000	
75%	0.000000			
max	104060.000000	max	104060.000000	
max	2200.000000			
Name: V337, dtype: float64		Name: V337, dtype: float64		
Name: V337, dtype: float64				

The unique values in V337 feature is: 255

V337 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V338 feature

Five Number Summary

Feature Description		Feature Description of legit transactions		Feature Description of fraud transactions
count	82351.000000	count	78665.000000	
count	3686.000000			
mean	151.160542	mean	155.748691	
mean	53.242268			
std	1095.034387	std	1119.429005	
std	190.250383			
min	0.000000	min	0.000000	
min	0.000000			
25%	0.000000	25%	0.000000	
25%	0.000000			
50%	0.000000	50%	0.000000	
50%	0.000000			

1. self_case_study_1_IEEE_fraud_detection_EDA

75%	0.000000	75%	0.000000
75%	0.000000		
max	104060.000000	max	104060.000000
max	2400.000000		
Name: V338, dtype: float64		Name: V338, dtype: float64	
Name: V338, dtype: float64			

The unique values in V338 feature is: 381
V338 feature has 508189 null values, which holds 0.8605496664070174 %

Analysis of V339 feature

Five Number Summary

Feature Description of legit transactions	Feature Description of fraud transactions	Feature Description
count 78665.000000	count 3686.000000	count 82351.000000
mean 103.287590	mean 45.496473	mean 100.700882
std 832.861349	std 175.966280	std 814.946722
min 0.000000	min 0.000000	min 0.000000
25% 0.000000	25% 0.000000	25% 0.000000
50% 0.000000	50% 0.000000	50% 0.000000
75% 0.000000	75% 0.000000	75% 0.000000
max 104060.000000	max 104060.000000	max 2400.000000
Name: V339, dtype: float64	Name: V339, dtype: float64	Name: V339, dtype: float64
Name: V339, dtype: float64		

The unique values in V339 feature is: 335
V339 feature has 508189 null values, which holds 0.8605496664070174 %

Observation:

1. Almost every V_feature has a value of 0 in majority of transactions.
2. Based on the missing value and number of unique value analysis we can find multiple V_features to have the same number of missing values and the same number of unique values. This pattern is widely present in the V_features.
3. In the first look these features does not look very important but we will see later on whether these are really useful or not.

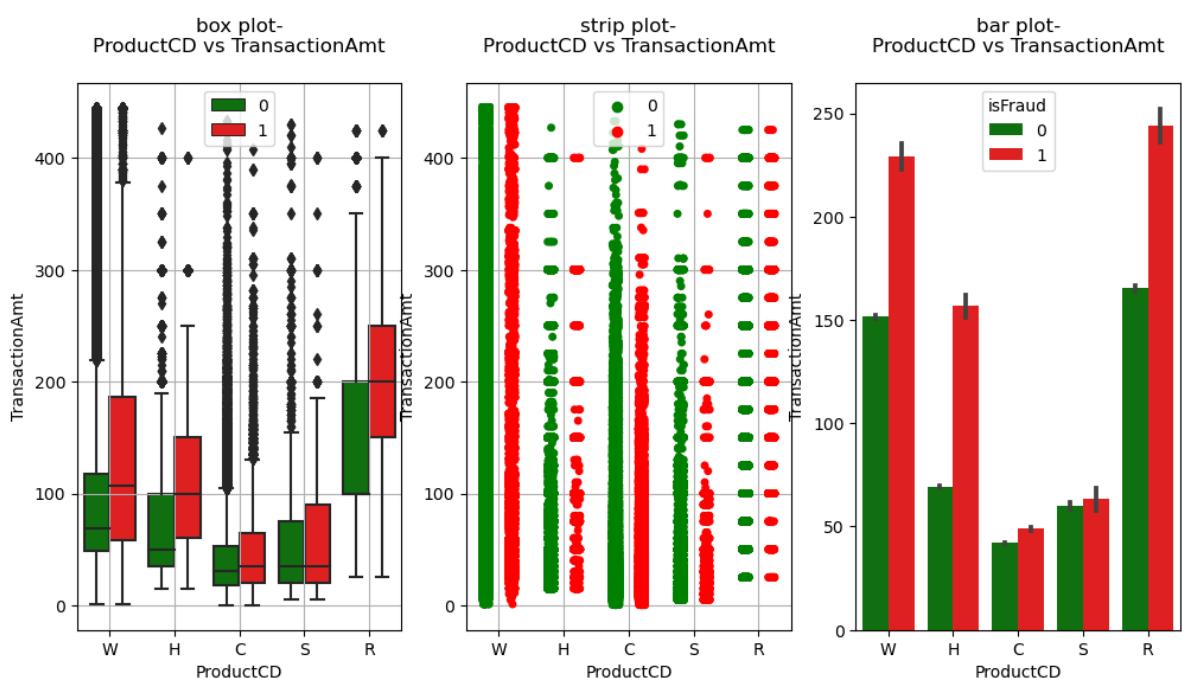
2.0 Bivariate Analysis

2.1 ProductCD V/S TransactionAmt

```
In [84]: plt.figure(figsize=(12,6))
plt.subplot(1,3,1)
sns.boxplot(x='ProductCD', y='TransactionAmt', data=df_train[df_train['TransactionAmt']>=np.percentile(df_train['TransactionAmt'], 99)])
plt.title(f'box plot-\nProductCD vs TransactionAmt\n')
plt.grid()
plt.legend()

plt.subplot(1,3,2)
df1 = df_train[df_train['TransactionAmt']<=np.percentile(df_train['TransactionAmt', 1)]
sns.stripplot(x = df1['ProductCD'], y = df1['TransactionAmt'], hue = df1['isFraud'])
plt.title(f'strip plot-\nProductCD vs TransactionAmt\n')
plt.grid()
plt.legend()

plt.subplot(1,3,3)
sns.barplot(x='ProductCD', y='TransactionAmt', hue ='isFraud', palette = ['g','r'])
plt.title(f'bar plot-\nProductCD vs TransactionAmt\n')
plt.show()
```



Observation:

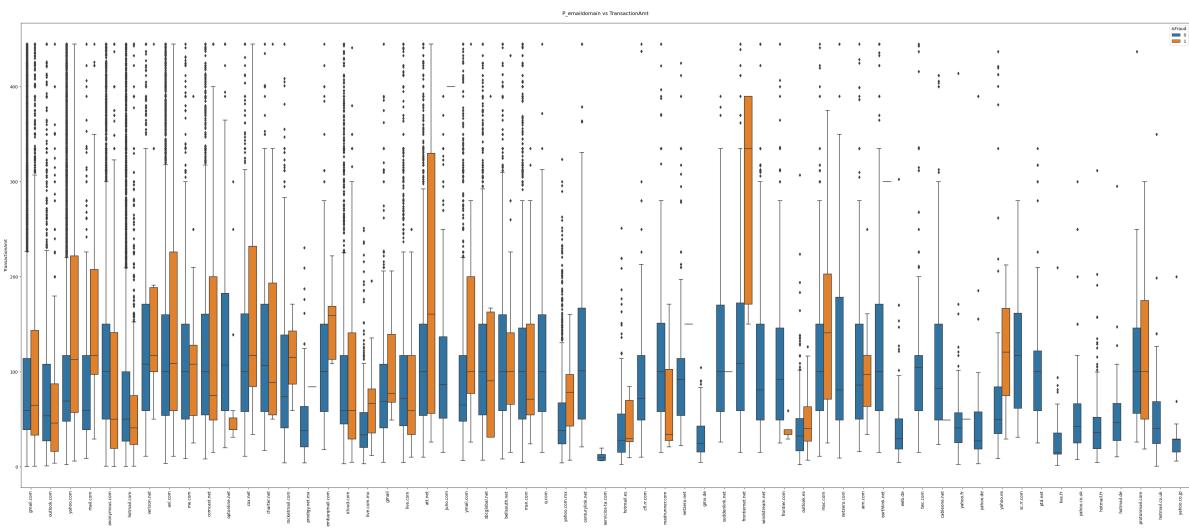
1. Product "R" has higher percentile values than the other products.
2. Despite the fact that roughly 6% of products fall into the "R" category. It has a larger transaction value than the "W" category, which happens 74% more frequently. This

indicates that compared to the other categories, the transaction volume in the product category "R" is way too high.

3. The distribution of fraudulent transactions is always higher than the distribution of legitimate transactions.

2.2 P_emaildomain V/S TransactionAmt

```
In [85]: plt.figure(figsize=(50,20))
ax = sns.boxplot(x='P_emaildomain', y='TransactionAmt', data=df_train[df_train['Type'] == 'Fraud'])
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha='center')
plt.title('\nP_emaildomain vs TransactionAmt\n')
plt.show()
```



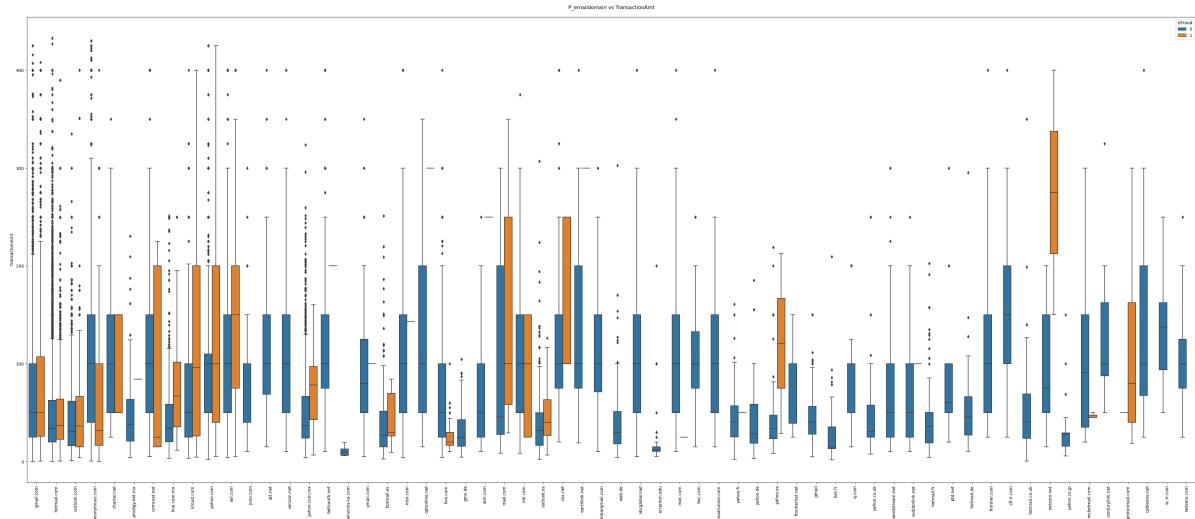
Observation:

1. As shown in the plot above, fraudulent transactions typically involve significant transaction amounts, with the exception of transactions done through Optonline.net, in which case the distribution of fraudulent transactions is less than that of legitimate transactions.
2. There is a distinct difference between transaction amounts for legitimate and fraudulent transactions for emails like yahoo.com, mail.com, ymail.com, frontimet.net, and yahoo.es.

2.3 R_emaildomain V/S TransactionAmt

```
In [86]: plt.figure(figsize=(50,20))
ax = sns.boxplot(x='R_emaildomain', y='TransactionAmt', data=df_train[df_train['Type'] == 'Fraud'])
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha='center')
plt.title('\nP_emaildomain vs TransactionAmt\n')
```

```
Out[86]: Text(0.5, 1.0, '\nP_emaildomain vs TransactionAmt\n')
```



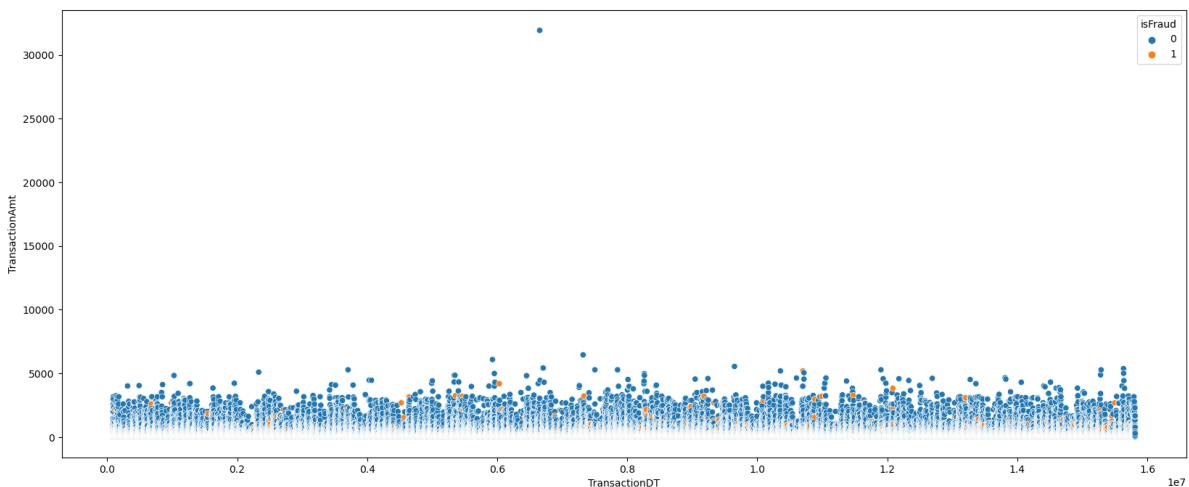
Observation:

1. For emails like netzero.net, yahoo.es, and cox.net, the transaction quantities for genuine and fraudulent transactions fluctuate noticeably.
2. The email address from which fraudulent transactions with the highest average value occurred was netzero.es.

2.4 TransactionDT V/S TransactionAmt

```
In [87]: plt.figure(figsize=(20,8))
sns.scatterplot(x = df_train['TransactionDT'] , y = df_train['TransactionAmt'] , hue = df_train['isFraud'])

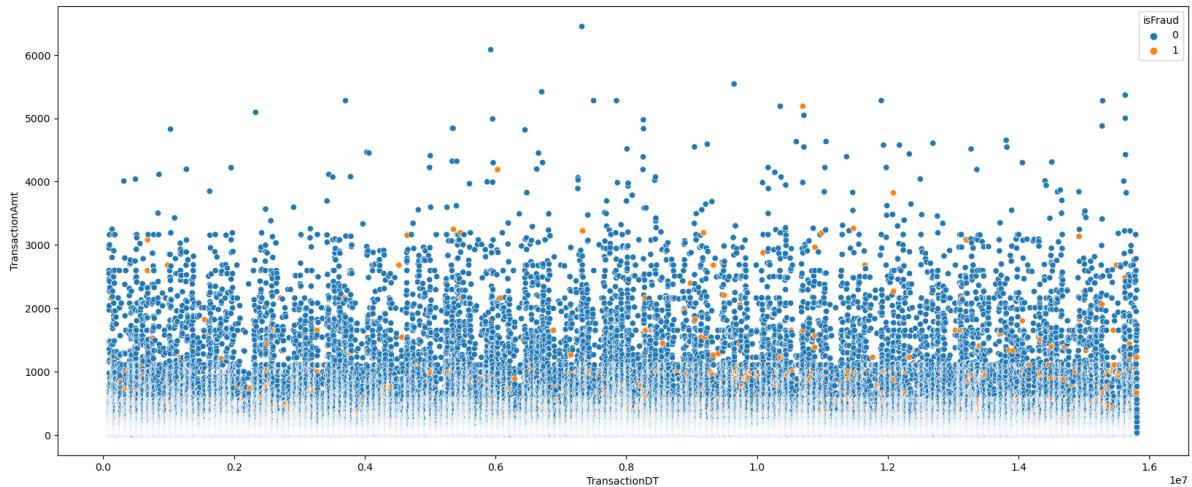
Out[87]: <Axes: xlabel='TransactionDT', ylabel='TransactionAmt'>
```



```
In [88]: # There is one extreme outlier in TransactionAmt
# We will remove that outlier
df_train = df_train[df_train['TransactionAmt']<30000]
```

```
In [89]: plt.figure(figsize=(20,8))
sns.scatterplot(x = df_train['TransactionDT'] , y = df_train['TransactionAmt'] , hue = df_train['isFraud'])

Out[89]: <Axes: xlabel='TransactionDT', ylabel='TransactionAmt'>
```

**Observation:**

1. The frequency of fraudulent transactions has increased along with the passing of time.

2.5 Card (1-6) V/S TransactionAmt

```
In [90]: card_features = ['card1', 'card2', 'card3', 'card4', 'card5', 'card6']

In [91]: def bivariate_plots(data, feature):
    if (data[feature].dtype == 'int64') or (data[feature].dtype == float):
        plt.figure(figsize=(20,8))
        sns.scatterplot(x = data[feature] , y = data['TransactionAmt'], hue = data[feature])
        plt.show()

    else:
        plt.figure(figsize=(12,6))
        plt.subplot(1,3,1)
        sns.boxplot(x=feature, y='TransactionAmt', data=data)
        plt.title(f'box plot-{feature} vs TransactionAmt')
        plt.xticks(rotation = 90)
        plt.grid()
        plt.legend()

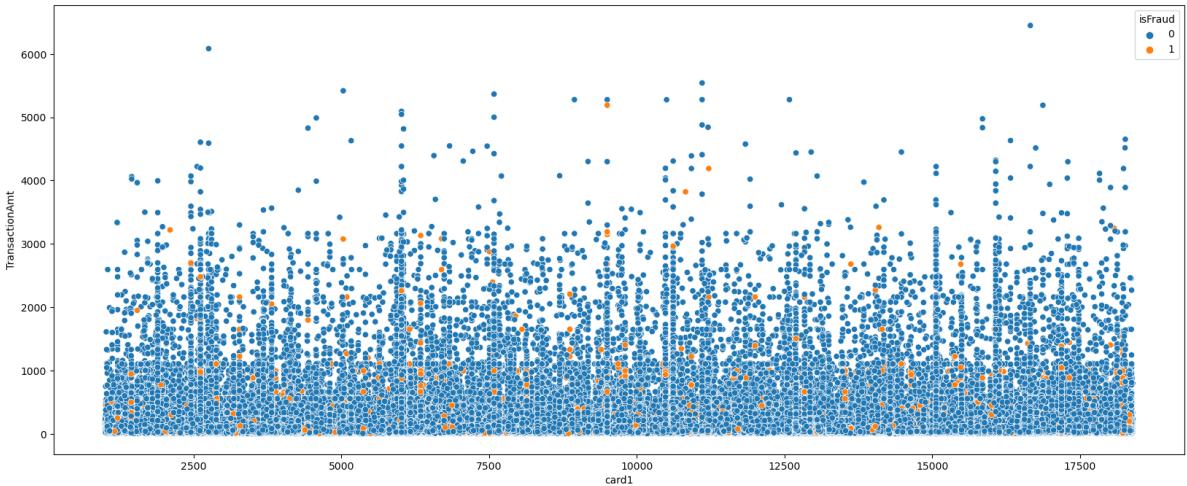
        plt.subplot(1,3,2)
        df1 = data[data['TransactionAmt'] <= np.percentile(data['TransactionAmt'], 99)]
        sns.stripplot(x = df1[feature], y = df1['TransactionAmt'], hue = df1['isFraud'])
        plt.title(f'strip plot-{feature} vs TransactionAmt')
        plt.xticks(rotation = 90)
        plt.grid()
        plt.legend()

        plt.subplot(1,3,3)
        sns.barplot(x=feature, y='TransactionAmt', hue ='isFraud', palette = ['green', 'red'])
        plt.title(f'bar plot-{feature} vs TransactionAmt')
        plt.xticks(rotation = 90)
        plt.show()

In [92]: for fea in card_features:
    print(f'Analysis of {fea} feature w.r.t TransactionAmt')
    bivariate_plots(df_train, fea)
    print('\n')
```

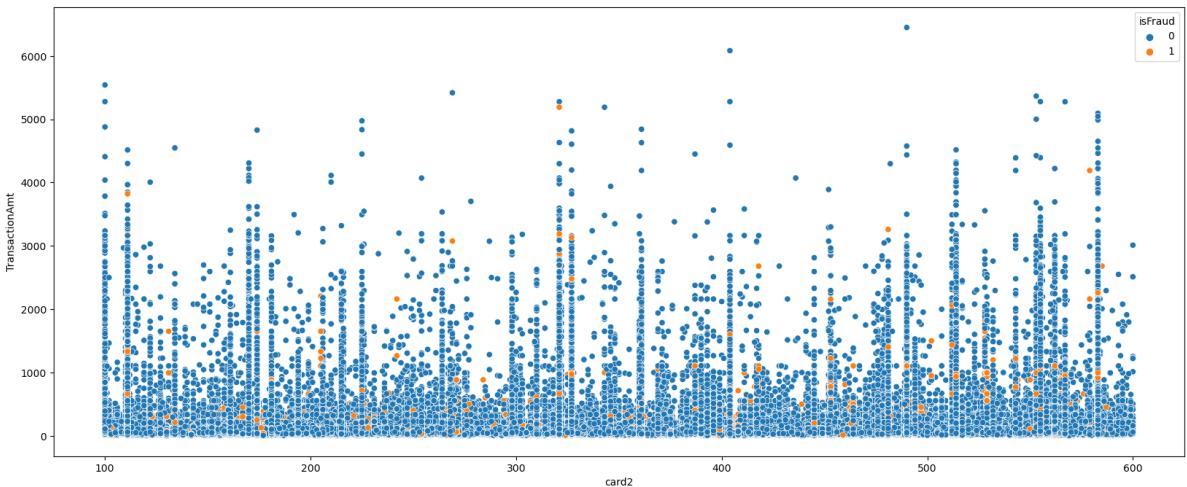
```
print('*'*120)
print('\n')
```

Analysis of card1 feature w.r.t TransactionAmt



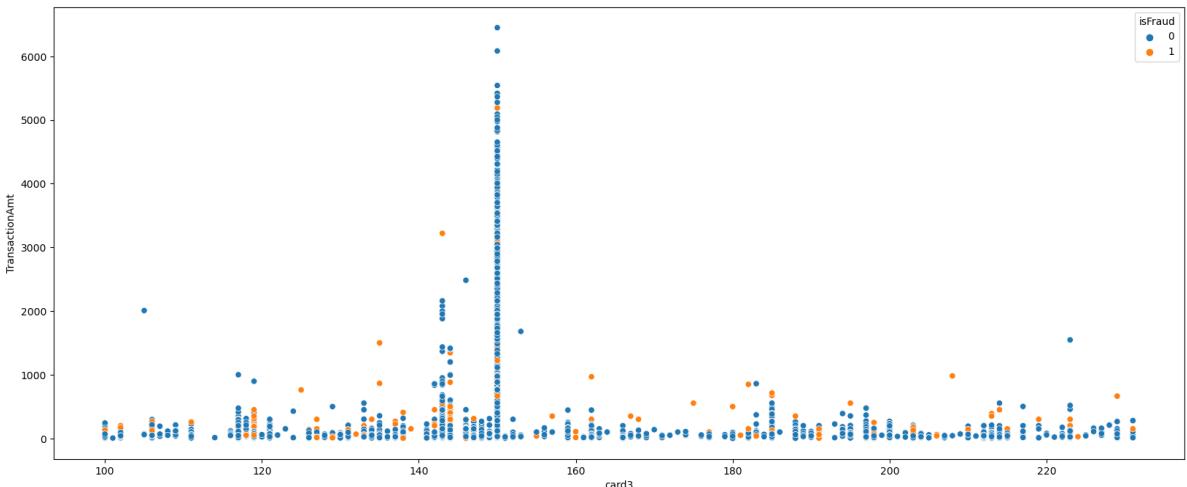
```
*****  
*****
```

Analysis of card2 feature w.r.t TransactionAmt

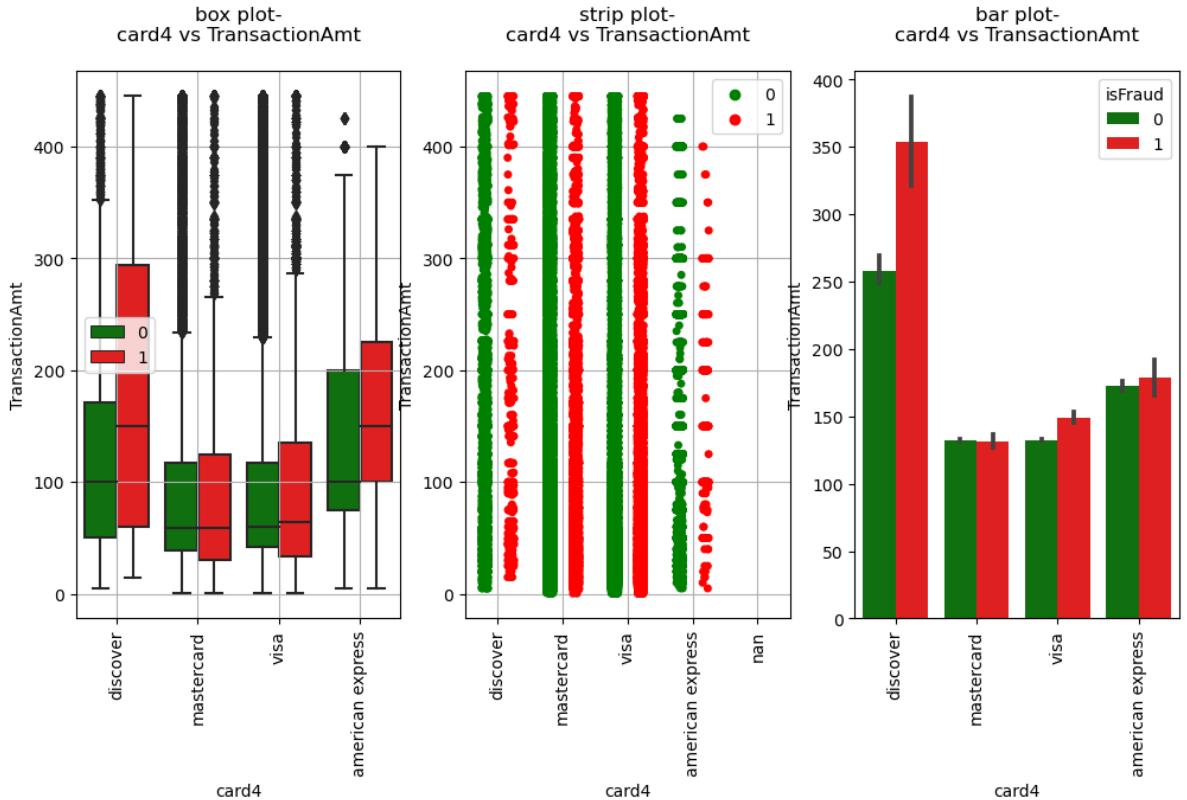


```
*****  
*****
```

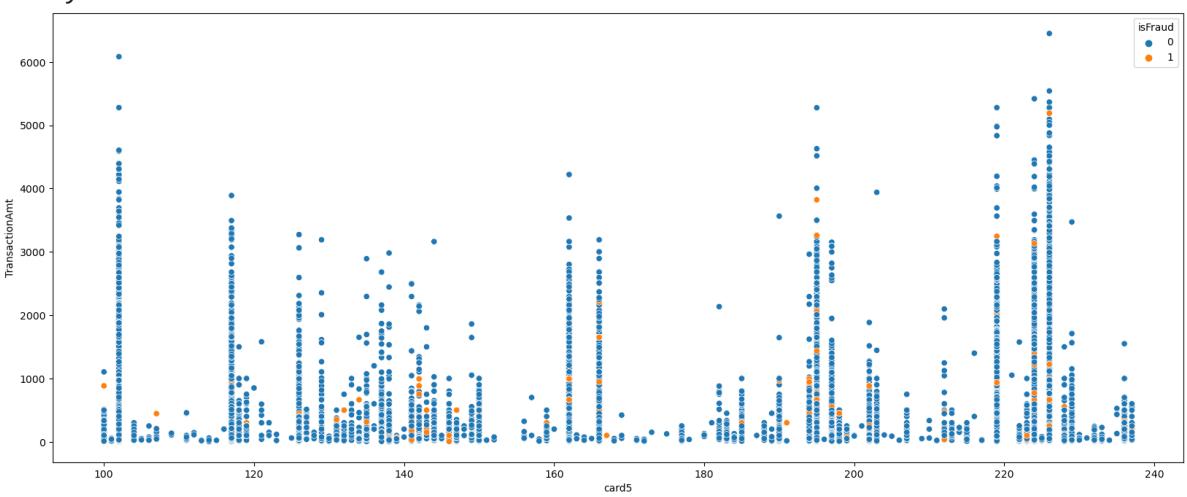
Analysis of card3 feature w.r.t TransactionAmt



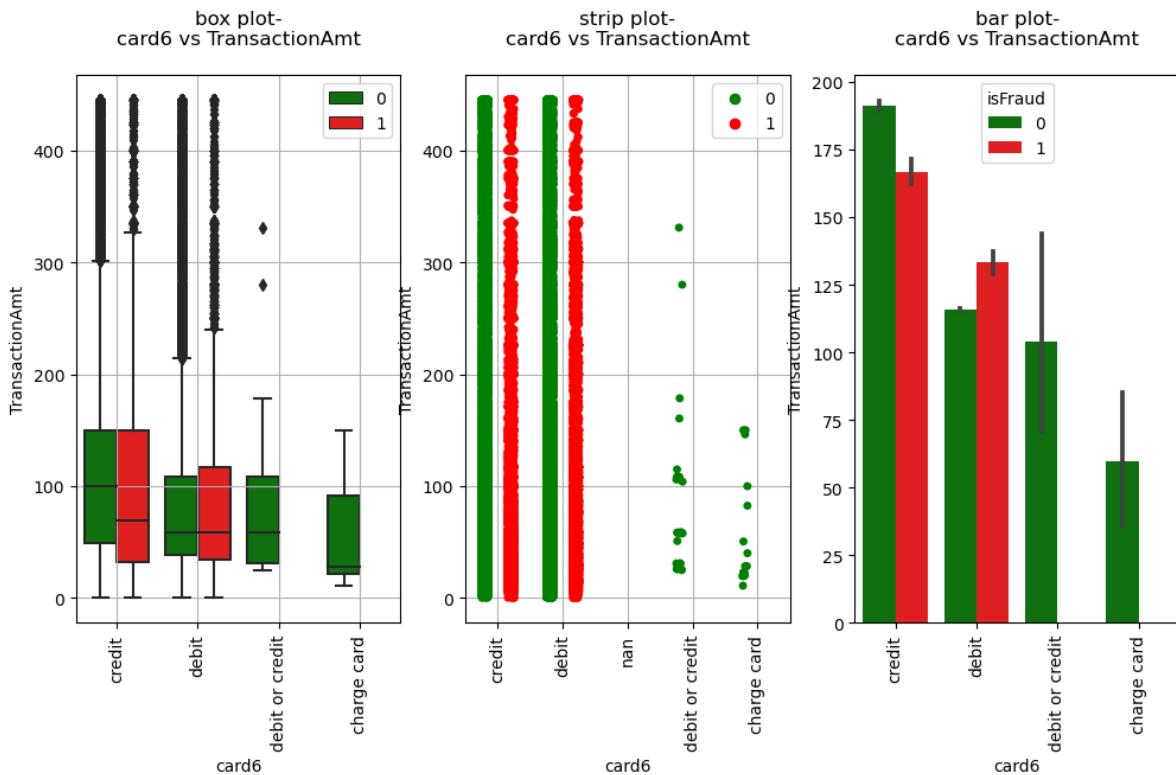
Analysis of card4 feature w.r.t TransactionAmt



Analysis of card5 feature w.r.t TransactionAmt



Analysis of card6 feature w.r.t TransactionAmt



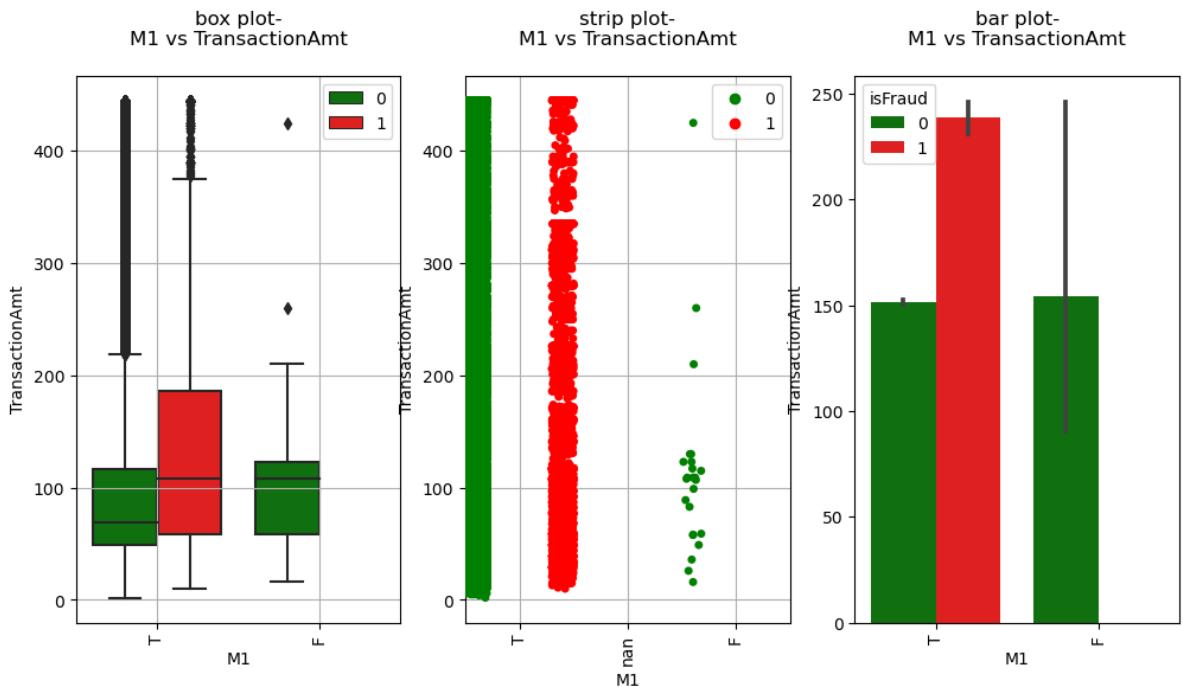
Observation:

1. Since we are aware that Card1, Card2, Card3 & Card5 are numerical categorical variables, we are unable to deduce much from their plots.
2. Card 4: According to a Univariate analysis, the category with the greatest number of fraud transactions is "discover." In this instance as well, the 'discover' category alone had the greatest average transaction value compared to the other categories for both legitimate and fraudulent transactions.
3. Card6: Those who make purchases with credit cards appear to spend more money. Also, When using a credit card, fraudulent transactions typically have a lower transaction amount than legitimate ones.

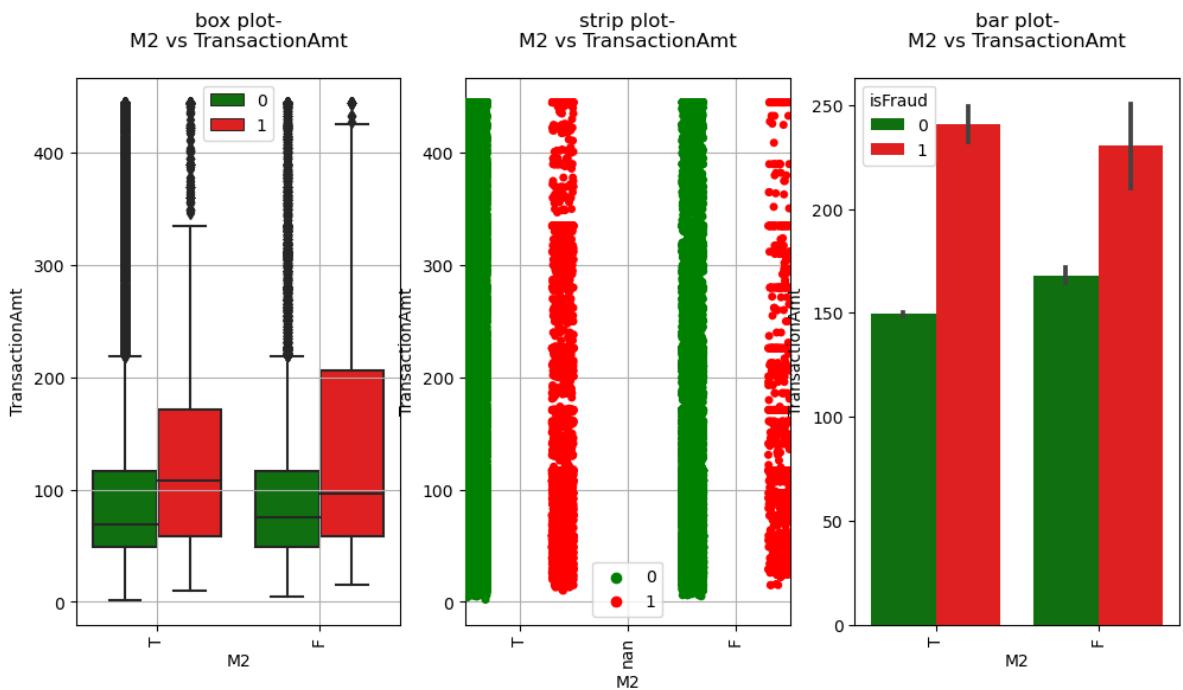
2.6 M (1-9) V/S TransactionAmt

```
In [93]: m_features = ['M1', 'M2', 'M3', 'M4', 'M5', 'M6', 'M7', 'M8', 'M9']
for fea in m_features:
    print(f'Analysis of {fea} feature w.r.t TransactionAmt')
    bivariate_plots(df_train, fea)
    print('\n')
    print('*'*120)
    print('\n')
```

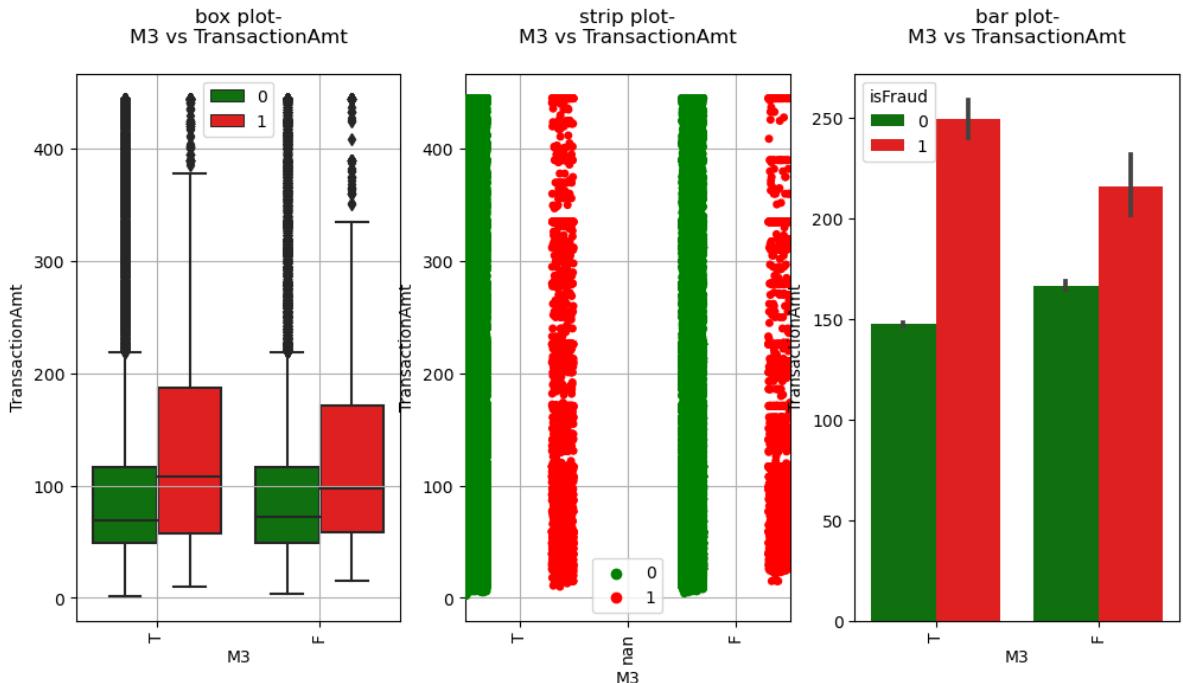
Analysis of M1 feature w.r.t TransactionAmt



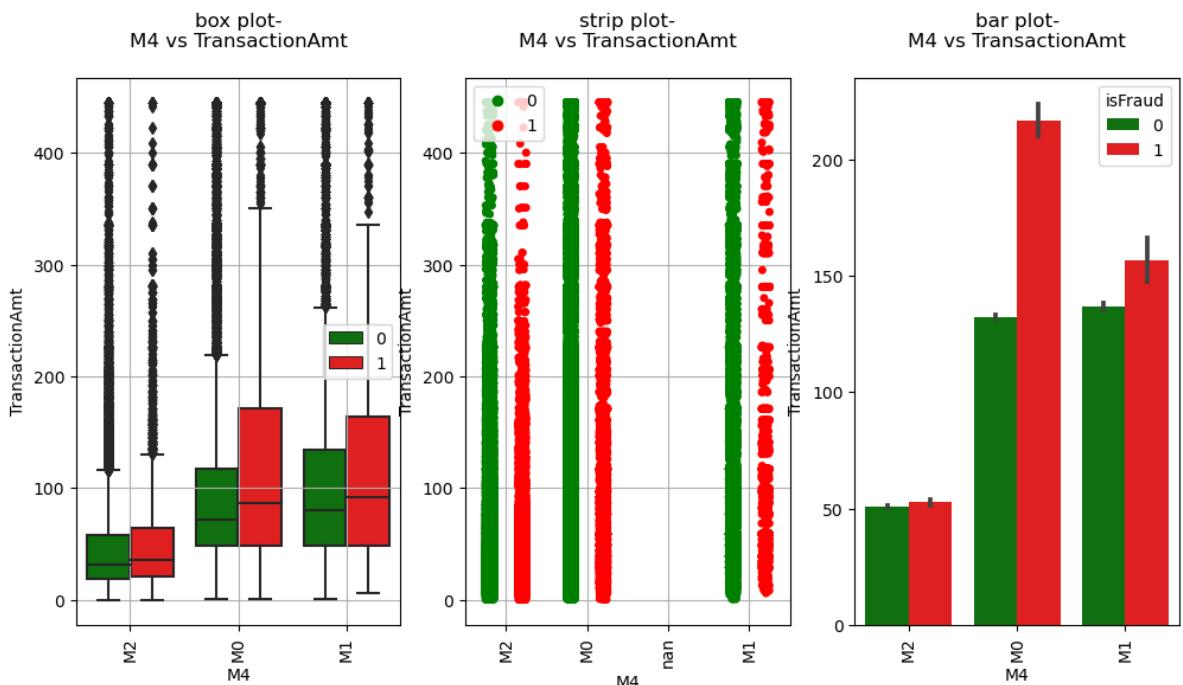
Analysis of M2 feature w.r.t TransactionAmt



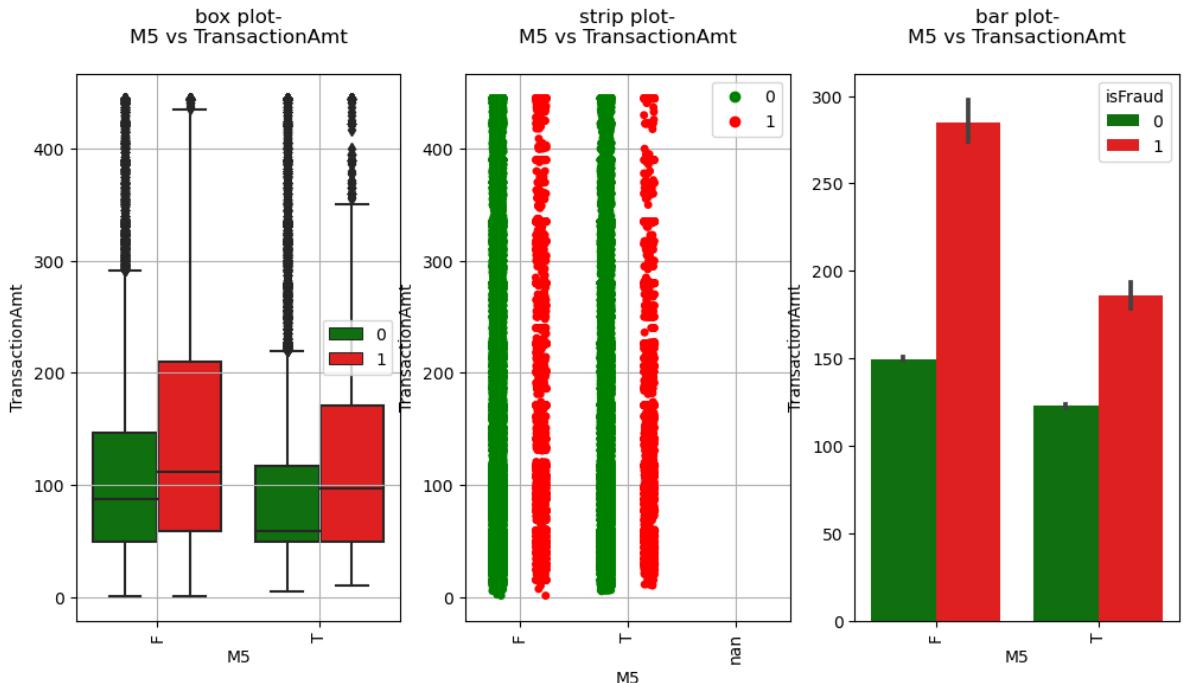
Analysis of M3 feature w.r.t TransactionAmt



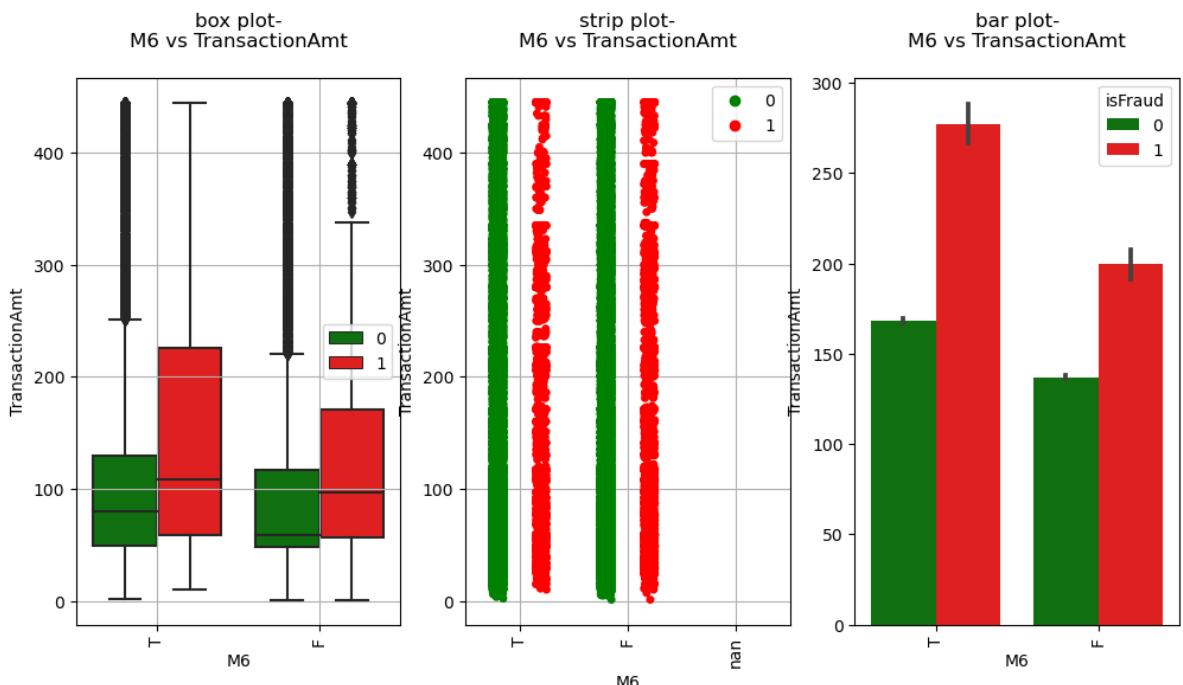
Analysis of M4 feature w.r.t TransactionAmt



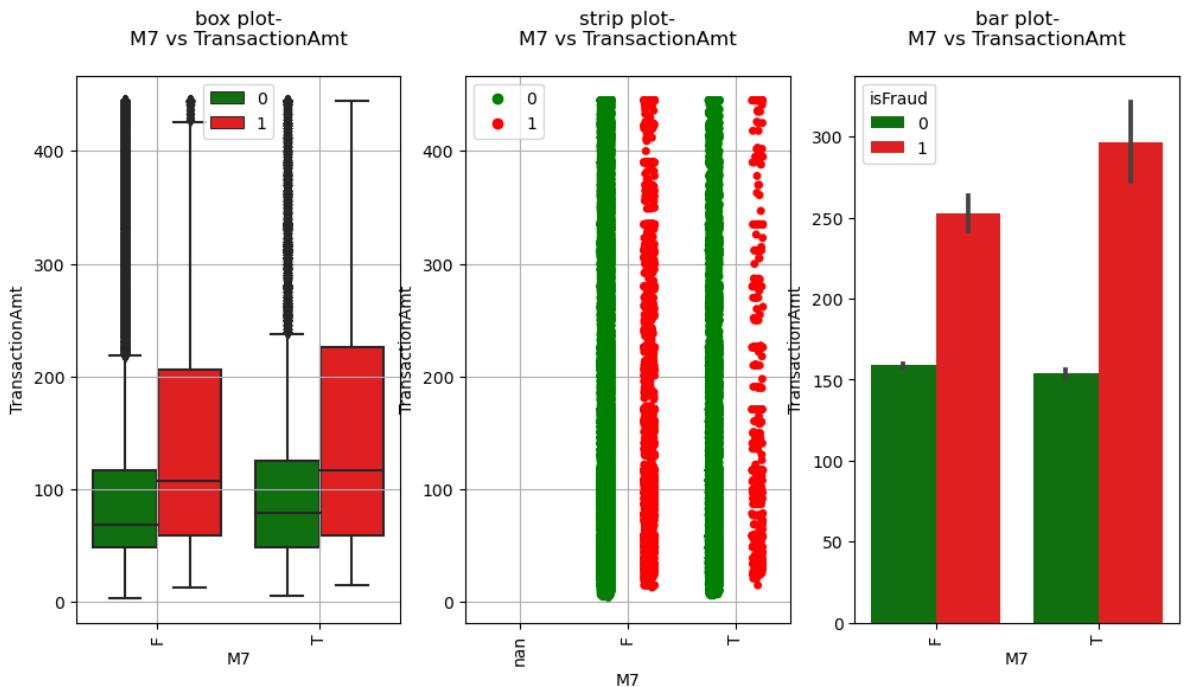
Analysis of M5 feature w.r.t TransactionAmt



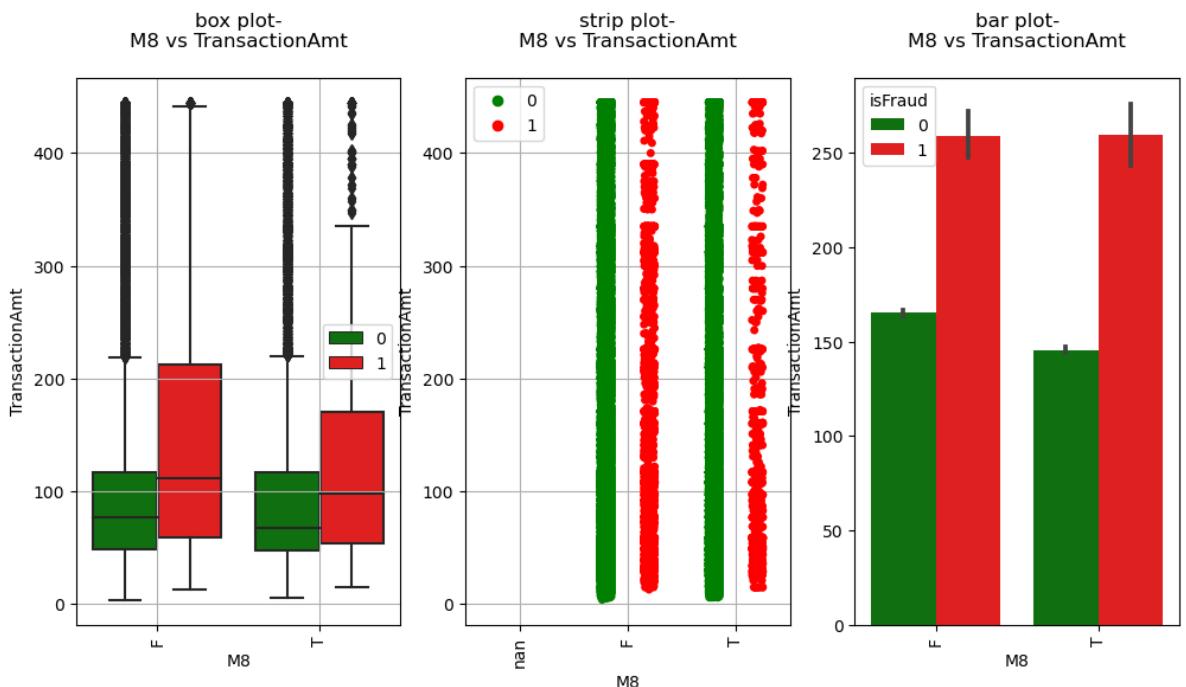
Analysis of M6 feature w.r.t TransactionAmt



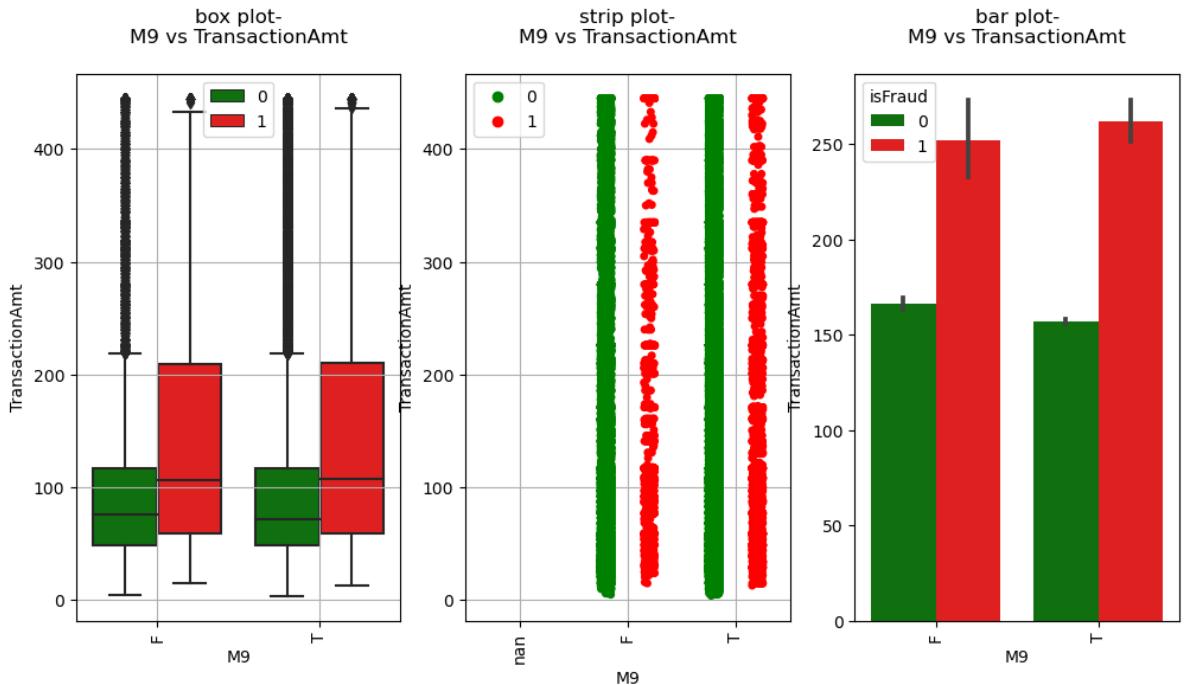
Analysis of M7 feature w.r.t TransactionAmt



Analysis of M8 feature w.r.t TransactionAmt



Analysis of M9 feature w.r.t TransactionAmt



Observation:

1. Across all categories, the amount of fraudulent transactions is generally higher than the amount of legit ones.

2.7 Device Features V/S TransactionAmt

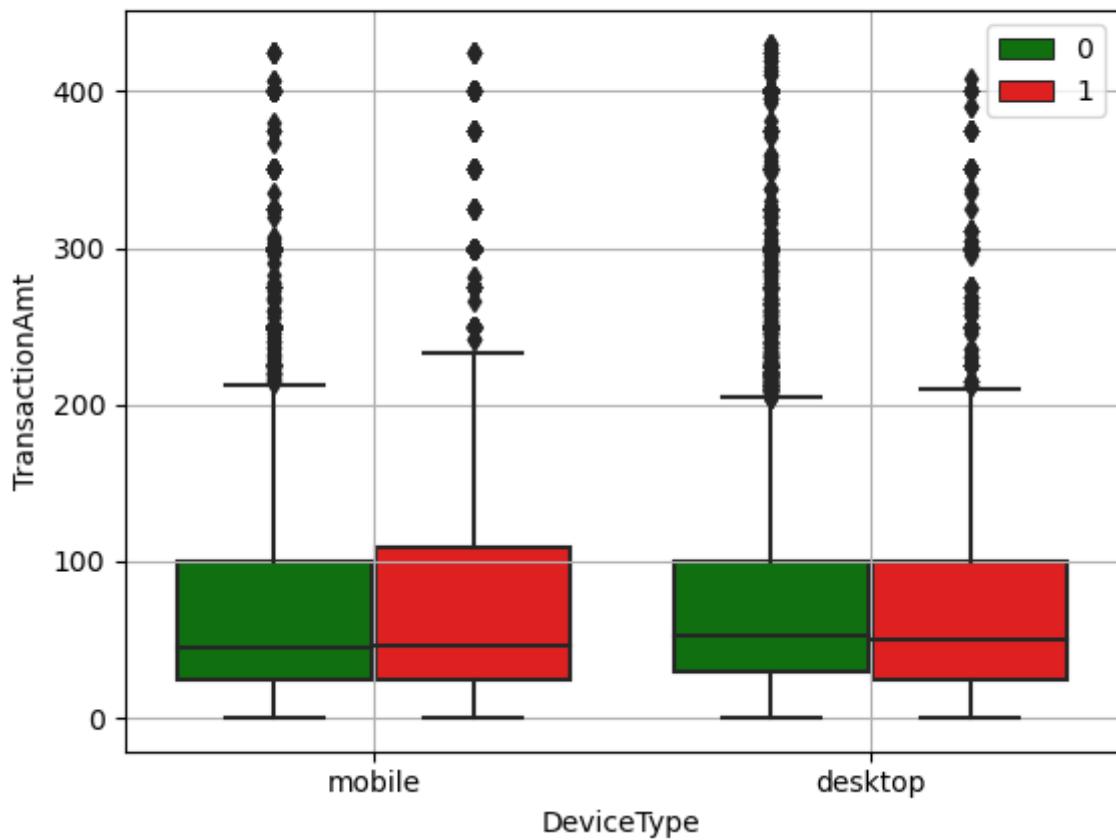
In [94]:

```
# DeviceType
sns.boxplot(x='DeviceType', y='TransactionAmt', data=df_train[df_train['TransactionAmt']<=np.percentile(df_train['TransactionAmt'], 99.5)])
plt.title(f'box plot-\nDeviceType vs TransactionAmt\n')
plt.grid()
plt.legend()
plt.show()

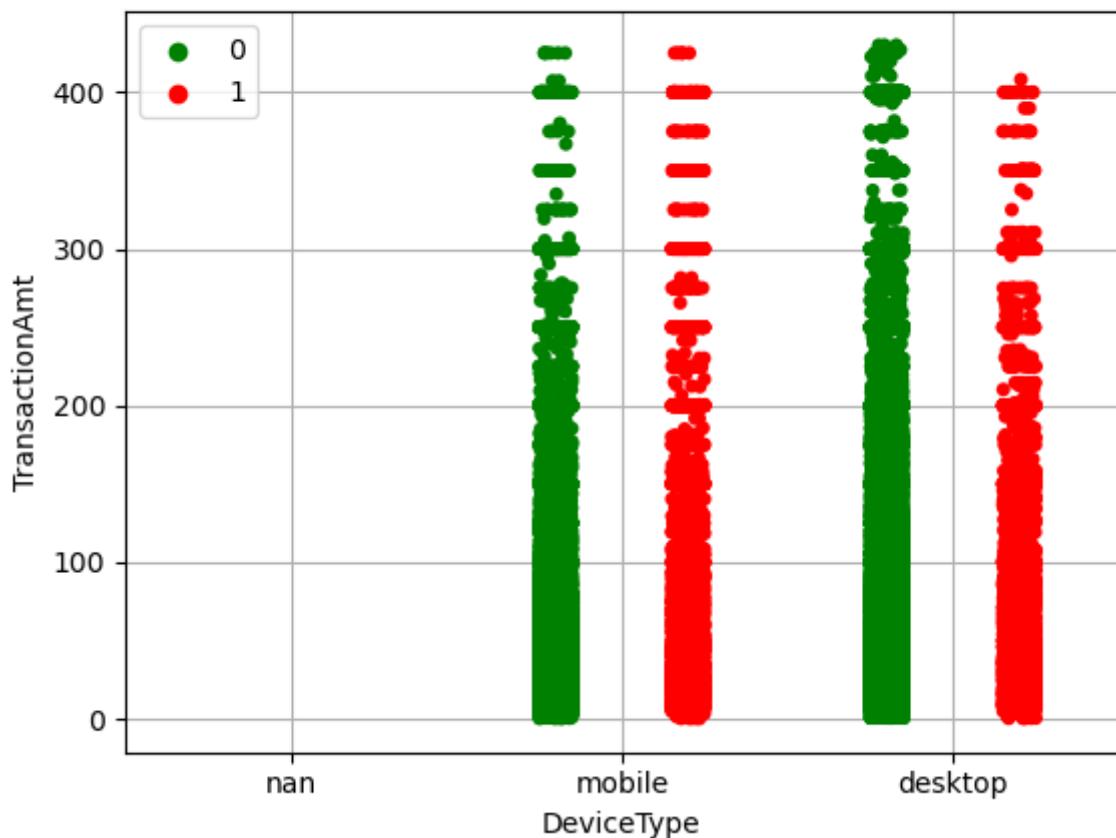
df1 = df_train[df_train['TransactionAmt']<=np.percentile(df_train['TransactionAmt'], 99.5)]
sns.stripplot(x = df1['DeviceType'], y = df1['TransactionAmt'], hue = df1['isFraud'])
plt.title(f'strip plot-\nDeviceType vs TransactionAmt\n')
plt.grid()
plt.legend()
plt.show()

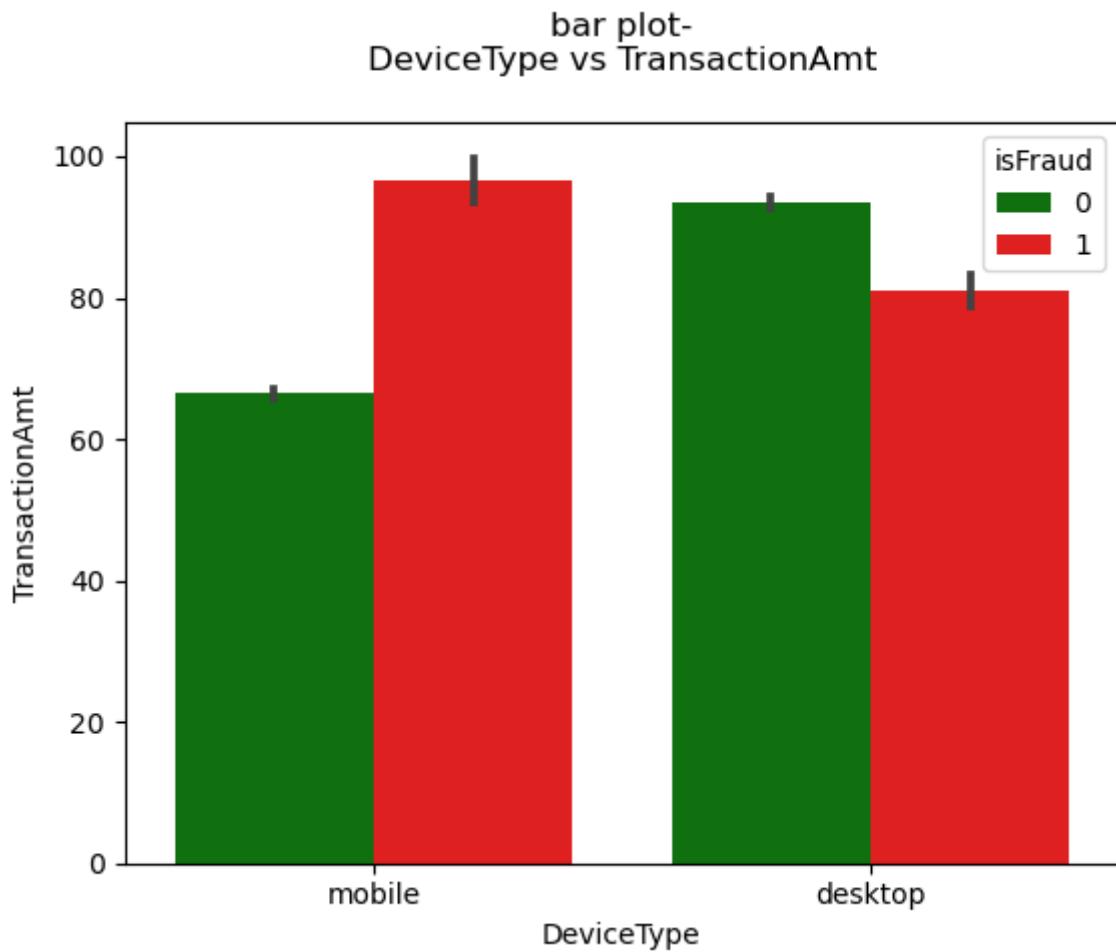
sns.barplot(x='DeviceType', y='TransactionAmt', hue ='isFraud', palette = [ 'g', 'r'])
plt.title(f'bar plot-\nDeviceType vs TransactionAmt\n')
plt.show()
```

box plot-
DeviceType vs TransactionAmt



strip plot-
DeviceType vs TransactionAmt



**Observation:**

1. Compared to mobile transactions, the average transaction amount for fraudulent transactions is lower on desktop than it is for genuine transactions.